

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2004, 09:34:43 ; Search time 142 Seconds
(without alignments)
2108.639 Million cell updates/sec

Title: US-09-671-687A-3
Perfect score: 5034
Sequence: 1 MSGSLMSQKVTSPYWEERI.....RLLCDAYMCWQSPYSLYK 949

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 segs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : \$PTREMBL_25.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_virus.*
 - 16: sp_bacteriaph.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4983	99.0	960	094934	Q94934 homo sapien
2	4971.5	98.8	956	Q9NQC7	Q9nqc7 homo sapien
3	4718.5	93.7	952	Q8CGB0	Q8cgb0 mus musculus
4	4718.5	93.7	959	Q80TQ2	Q80tq2 mus musculus
5	2952	58.6	623	Q8BYL9	Q8byl9 mus musculus
6	2512	49.9	515	Q9NZX9	Q9nzx9 homo sapien
7	1662	33.0	313	Q80VB3	Q80vb3 mus musculus
8	1458.5	29.0	318	Q8BXZ3	Q8bxz3 mus musculus
9	1136.5	22.6	551	Q8IPC3	Q8ipc3 drosophila
10	1136.5	22.6	639	Q8IPC5	Q8ipc5 drosophila
11	1134	22.5	550	Q8IPC4	Q8ipc4 drosophila
12	1132.5	22.5	639	Q8SYF0	Q8syf0 drosophila
13	856	17.0	517	Q9VL04	Q9vl04 drosophila
14	548	10.9	970	Q9U3F9	Q9u3f9 caenorhabdi
15	149.5	3.0	1046	Q55156	Q55156 rattus norv
16	149	3.0	1046	Q9Z0H8	Q9z0h8 mus musculus

17	147.5	2.9	449	4	Q86WU4	Q86wu4 homo sapien
18	145	2.9	1556	5	Q9VRP1	Q9vrp1 drosophila
19	143.5	2.9	1047	11	Q9EP81	Q9ep81 mus musculus
20	143	2.8	547	4	Q8WML1	Q8wml1 homo sapien
21	142.5	2.8	429	4	Q96DZ5	Q96dz5 homo sapien
22	142	2.8	1012	11	Q7TSI9	Q7tsi9 mus musculus
23	141	2.8	1012	11	Q8CHU1	Q8chul mus musculus
24	140.5	2.8	429	4	Q96C99	Q96c99 homo sapien
25	140	2.8	924	16	Q7VDY2	Q7vdy2 prochloroco
26	139.5	2.8	1082	10	Q94DZ5	Q94dz5 oryza sativ
27	139.5	2.8	1921	5	Q86BR0	Q86br0 drosophila
28	138	2.7	724	11	Q8BWD1	Q8bwd1 mus musculus
29	138	2.7	1024	4	Q14527	Q14527 homo sapien
30	138	2.7	1391	11	Q922J3	Q922j3 mus musculus
31	137	2.7	868	10	Q9SDN6	Q9sdn6 nicotiana t
32	135.5	2.7	2273	16	Q31152	Q31152 neisseria m
33	135	2.7	1046	4	Q9UDT6	Q9udt6 homo sapien
34	134.5	2.7	657	4	Q7Z3N8	Q7z3n8 homo sapien
35	132.5	2.6	419	4	Q43611	Q43611 homo sapien
36	131	2.6	2042	10	Q9M9P8	Q9m9p8 arabidopsis
37	129.5	2.6	832	5	Q8LMZ4	Q8lmz4 oryza sativ
38	128.5	2.6	634	5	Q06279	Q06279 schistosoma
39	128.5	2.6	840	11	Q8C0S5	Q8c0s5 mus musculus
40	128	2.5	560	3	Q874X2	Q874x2 podospira a
41	127	2.5	803	10	Q8H7I2	Q8h7i2 oryza sativ
42	127	2.5	804	10	Q8S7A2	Q8s7a2 oryza sativ
43	127	2.5	804	10	Q7XFN0	Q7xfn0 oryza sativ
44	126.5	2.5	856	5	Q9V4V6	Q9v4v6 drosophila
45	125.5	2.5	624	11	Q7TPM8	Q7tpm8 mus musculus

ALIGNMENTS

RESULT 1
Q94934
AC Q94934 Q96EH0 PRELIMINARY; PRT; 960 AA.
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE Hypothetical protein KIAA0849 (Cylinclromatosis) (Turban tumor syndrome) (Fragment).
GN KIAA0849
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K., Koyama M., Kikuno R., Hirose M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";
RL DNA Res. 5:355-364(1998)
RN [2]
RP SEQUENCE OF 8-960 FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB020656; BAA74872.2;
DR EMBL; BC012342; AAH12342.1;
DR Genew; HGNC:2584; CYLD.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR InterPro; IPR000938; CAP-Gly.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF01302; CAP_GLY; 2.
DR Pfam; PF00443; UCH; 1.


```
Db 421 FSLTKMPTNGSGHSPUSLSAQSVMEELNTPVQESPPPLAMPNGSHGLEVSLEA 480
Qy 474 NPPFYGVIRWIGOPPGPLNEVLAGELEDEACAGCTDGTFRGTRFTYTCALKKALFVKL 533
Db 481 NPPFYGVIRWIGOPPGPLNEVLAGELEDEACAGCTDGTFRGTRFTYTCALKKALFVKL 540
Qy 534 PDSRFASLQVSNQIERCNSLAFGGYLSVVEENTPMPMEKEGLEIMIGKKKGIOGHYNS 593
Db 541 PDSRFASLQVSNQIERCNSLAFGGYLSVVEENTPMPMEKEGLEIMIGKKKGIOGHYNS 600
Qy 594 CYLDSTLFCFLAFSSVLDTVLLRPKEKNDVEYSETQELLRTETVNPRLRYGYVCATKIM 653
Db 601 CYLDSTLFCFLAFSSVLDTVLLRPKEKNDVEYSETQELLRTETVNPRLRYGYVCATKIM 660
Qy 654 KLRKILEKVEAASGFTSEEDKDPPEFLNLPFHILRPVPLKIRSAQKVDQCFYQIFME 713
Db 661 KLRKILEKVEAASGFTSEEDKDPPEFLNLPFHILRPVPLKIRSAQKVDQCFYQIFME 720
Qy 714 KNEKVGPTTQQLEWSFINSNLKFAEAPSCLIIQMPRFQKFLPKKIFPSLELNITDL 773
Db 721 KNEKVGPTTQQLEWSFINSNLKFAEAPSCLIIQMPRFQKFLPKKIFPSLELNITDL 780
Qy 774 LEDTPQCRICGGLAMEYECREYDDPDISAGKIKQFCKTCTNTQVHLHPKRLNHNKYNPVS 833
Db 781 LEDTPQCRICGGLAMEYECREYDDPDISAGKIKQFCKTCTNTQVHLHPKRLNHNKYNPVS 840
Qy 834 PKDLPDWDRHGICPQNMELFAVLCIETSHYVAFVKYKDDSAWLPFFDSMAORDGQNG 893
Db 841 PKDLPDWDRHGICPQNMELFAVLCIETSHYVAFVKYKDDSAWLPFFDSMAORDGQNG 900
Qy 894 FNIPQVTPCEVGEYLKMSLEDLHSLDSRRIOGCARRLLCDAYMCYQSTMSLYK 949
Db 901 FNIPQVTPCEVGEYLKMSLEDLHSLDSRRIOGCARRLLCDAYMCYQSTMSLYK 956

RESULT 3
Q8CGB0 PRELIMINARY; PRT; 952 AA.
AC Q8CGB0;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DE 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Similar to cylindromatosis (turban tumor syndrome).
GN 2010013M14RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042438; AAH42438.1; -.
DR MGD; MGI:1921506; 2010013M14RIK.
DR GO; GO:0004197; F.cysteine-type endopeptidase activity; IEA.
DR GO; GO:0004221; Fubiquitin thiolesterase activity; IEA.
DR GO; GO:0006511; Fubiquitin-dependent protein catabolism; IEA.
DR InterPro; IPR000938; CAP-Gly.
DR InterPro; IPR001394; Peptidase C19.
DR InterPro; IPR000408; Reg_chromatid_condens.
DR Pfam; PF01302; CAP_GLY; 2.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS0245; CAP_GLY 2; 2.
DR PROSITE; PS00626; RCL 2; 1.
DR PROSITE; PS0235; UCH 2.3; 1.
SQ SEQUENCE 952 AA; 106585 MW; 0AC0C7D4FF215A9C CRC64;
```

Query Match 93.7%; Score 4718.5; DB 11; Length 952;
Best Local Similarity 94.2%; Pred. No. 0;
Matches 898; Conservative 22; Mismatches 28; Indels 5; Gaps 5;

```
Qy 1 MSSGLWSQEKVTSYWEERIFYLLOECVTDKQTKLLKVPKSGISGYIQDRSVGHGR 60
Db 1 MSSGLWSQEKVTSYWEERIFYLLOECVTDKQTKLLKVPKSGISGYIQDRSVGHGR 60
Qy 61 PSAGKKNQIOLKLEOPHVLVDE-DVVEINEKFTTELLAITNCEERFSLFKNRRLS 119
Db 61 PSTKGNQKQIOLKLEOPHVLVDEKDVVEINEKFTTELLAITNCEERFSLFKNRRLS 120
Qy 120 KGLQIDVGCVPKQVLRSGEEKFPGVWRFRGPLLAERTVSGIFFGVVLEEBEGRGQFTDGV 179
Db 121 KGLQVGVSPKQVLRSGEEKFPGVWRFRGPLLAERTVSGIFFGVVLEEBEGRGQFTDGV 180
Qy 180 YQKQLFCQDEDCG-FVALDKLEIEDDDTALESIDYAGPDTMQVLPPLPINSRVLKG 238
Db 181 YQKQLFCQDEDCGVFVALDKLEIEDDDNGLSDFAFGPDTMQVLPPLPINSRVLKV 240
Qy 239 GETIESGTVIFCDVLPKESLGYFVGVDMDNPIGNWDRFGV-LCSFACVESTILLHN 297
Db 241 GESTESGTVIFCDVLPKESLGYFVGVDMDNPIGNWDRFGVQLCSFASVESTILLHN 300
Qy 298 DIIPESVTQERRPPKLAFLMSRGVGDGSSSHNKPKATGSTSDPGR-RSELYTLNGSSV 356
Db 301 DIIPDSVTQERRPPKLAFLMSRGVGDGSSSHNKPKVGTGSTSDPGRNRSELYTLNGSSV 360
Qy 357 DSQPSQKNTWYIDEVAEDPAKSLTEISTDFDRSSPPLOPPVNSLTENRHSPLFSL 416
Db 361 DSQ-QSKNPNWYIDEVAEDPAKSLTEMSDFGSHSPPPQPPSMNLSSENRFHSPLFSL 419
Qy 417 TKMPTNGSGHSPUSLSAQSVMEELNTPVQESPPPLAMPNGSHGLEVSLEAENPP 476
Db 420 TKMPTNGSGHSPUSLSAQSVMEELNTPVQESPPPLAMPNGSHGLEVSLEAENPP 479
Qy 477 FYGVIRWIGOPPGPLNEVLAGELEDEACAGCTDGTFRGTRFTYTCALKKALFVKL 536
Db 480 FYGVIRWIGOPPGPLNEVLAGELEDEACAGCTDGTFRGTRFTYTCALKKALFVKL 539
Qy 537 RFASLQVSNQIERCNSLAFGGYLSVVEENTPMPMEKEGLEIMIGKKKGIOGHYNSCYL 596
Db 540 RFASLQVSNQIERCNSLAFGGYLSVVEENTPMPMEKEGLEIMIGKKKGIOGHYNSCYL 599
Qy 597 DSTLFCFLAFSSVLDTVLLRPKEKNDVEYSETQELLRTETVNPRLRYGYVCATKIMKLR 656
Db 600 DSTLFCFLAFSSVLDTVLLRPKEKNDVEYSETQELLRTETVNPRLRYGYVCATKIMKLR 659
Qy 657 KILEKVEAASGFTSEEDKDPPEFLNLPFHILRPVPLKIRSAQKVDQCFYQIFMEKNE 716
Db 660 KILEKVEAASGFTSEEDKDPPEFLNLPFHILRPVPLKIRSAQKVDQCFYQIFMEKNE 719
Qy 717 KVGVPITQQLEWSFINSNLKFAEAPSCLIIQMPRFQKFLPKKIFPSLELNITDLED 776
Db 720 KVGVPITQQLEWSFINSNLKFAEAPSCLIIQMPRFQKFLPKKIFPSLELNITDLED 779
Qy 777 TPRQCRICGGLAMEYECREYDDPDISAGKIKQFCKTCTNTQVHLHPKRLNHNKYNPVS 836
Db 780 TPRQCRICGGLAMEYECREYDDPDISAGKIKQFCKTCTNTQVHLHPKRLNHNKYNPVS 839
Qy 837 LPDMDWRHGICPQNMELFAVLCIETSHYVAFVKYKDDSAWLPFFDSMAORDGQNGPNI 896
Db 840 LPDMDWRHGICPQNMELFAVLCIETSHYVAFVKYKDDSAWLPFFDSMAORDGQNGPNI 899
Qy 897 PQVTPCEVGEYLKMSLEDLHSLDSRRIOGCARRLLCDAYMCYQSTMSLYK 949
Db 900 PQVTPCEVGEYLKMSLEDLHSLDSRRIOGCARRLLCDAYMCYQSTMSLYK 952

RESULT 4
Q80TQ2 PRELIMINARY; PRT; 999 AA.
AC Q80TQ2;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE MKIAA0849 protein (Fragment).
```



```
Db 121 KGLQVDVGPVKYQLRSGBKFGCVVFRGRLAERTVSGIFFGVLELBERGGQFTDGV 180
Qy 180 YQKQLFQCCDDCG-FVALDKLELIEDDDTALESYAGPBDTMQVELPPLPPEINSRVSLKG 238
Db 181 YQKQLFQCCDDCGFVALDKLELIEDDDNGLESDFAGPBDTMQVEPPPLPPEINSRVSLKV 240
Qy 239 GETIESGTWIFCDVLPKESISLGFVGVDMNDPIGNWMDGRFDGV-LCSFACVESTILLHIN 297
Db 241 GESTESGTWIFCDVLPKESISLGFVGVDMNDPIGNWMDGRFDGVLCFSFASVESTILLHIN 300
Qy 298 DIIP---ESVTQRRPPKLAFLMRSGVDGKSSSHNPKATGSTDPCNR-RSELYFTLNG 353
Db 301 DIIPALSDSVTQRRPPKLAFLMRSGVDGKSSSHNPKATGSTDPCNRSELYFTLNG 360
Qy 354 SSVDSQPSQSKNTWIDEVAEDPAKSLTEISDFDRSSPPLQPPVNSLTENRFSHLP 413
Db 361 SSVDSQ-QSKSKNPWIDEVAEDPAKSLTEISDFGHSSPPQPPVNSLSSENRFHSLP 419
Qy 414 FSLTKPNTNGSIGHSPSLSAQSVMEELNAPVQESPPPLAMPNGSHGLEVGSLAEVKE 473
Db 420 FSLTKPNTNGSMAHSPSLSAQSVMEELNSTPVQESPPPLPISSGNAHGLEVGSLAEVKE 479
Qy 474 NPFFYGIWIRGPGVNEVLGLELEDEACAGCTDGTFRGTRYFTCALKKALFVKLXSCR 533
Db 480 NPFFYGIWIRGPPGLSDVLAGLELEDEACAGCTDGTFRGTRYFTCALKKALFVKLXSCR 539
Qy 534 PDSRFASLPQVSNQIERCNSLAFGGYLSEVVEENTPPKMEKEGLEIMIGKKGIQGHYNS 593
Db 540 PDSRFASLPQVSNQIERCNSLAFGGYLSEVVEENTPPKMEKEGLEIMIGKKGIQGHYNS 599
Qy 594 CYLDSTLFLCLFAPSSVLDTVLLRP 617
Db 600 CYLDSTLFLCLFAPSSALDTVLLRP 623
```

RESULT 6

```
Q9NZX9 PRELIMINARY; PRT; 515 AA.
ID Q9NZX9 AC Q9NZX9
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE HSPC057.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=20499367; PubMed=11042152;
RA Zhang Q.H., Ye M., Wu X.Y., Ren S.X., Zhao M., Zhao C.J., Pu G.,
RA Shen Y., Fan H.Y., Lu G., Zhong M., Xu X.R., Han Z.G., Zhang J.W.,
RA Tao J., Huang Q.H., Zhou J., Hu G.X., Gu J., Chen S.J., Chen Z.;
RT "Cloning and functional analysis of cDNAs with open reading frames for
RT 300 previously undefined genes expressed in CD34+ hematopoietic
RT stem/progenitor cells.";
RL Genome Res. 10:1546-1560 (2000).
DR EMBL; AF161542; AAF29029.1; -.
DR InterPro; IPR000938; CAP-Gly.
DR Pfam; PF01302; CAP_Gly; 1.
DR PROSITE; P550245; CAP_GLY_2; 1.
SQ SEQUENCE 515 AA; 58197 MW; 23D82D7BD66EA146 CRC64;
```

Query Match 49.9%; Score 2512; DB 4; Length 515;
Best Local Similarity 94.3%; Pred. No. 2.6e-189;
Matches 477; Conservative 1; Mismatches 28; Indels 0; Gaps 0;

```
Qy 419 MENTNGSIGHSPSLSAQSVMEELNAPVQESPPPLAMPNGSHGLEVGSLAEVKEPNPPFY 478
Db 1 MENTNGSIGHSPSLSAQSVMEELNAPVQESPPPLAMPNGSHGLEVGSLAEVKEPNPPFY 60
Qy 479 GVIWIRGPGVNEVLGLELEDEACAGCTDGTFRGTRYFTCALKKALFVKLXSCRPSDRF 538
```

```
Db 61 GVIRWIGQPPGLNEVLGLELEDEACAGCTDGTFRGTRYFTCALKKALFVKLXSCRPSDRF 120
Qy 539 ASLQPVSNQIERCNSLAFGGYLSEVVEENTPPKMEKEGLEIMIGKKGIQGHYNSCYLDS 598
Db 121 ASLQPVSNQIERCNSLAFGGYLSEVVEENTPPKMEKEGLEIMIGKKGIQGHYNSCYLDS 180
Qy 599 TLFCLFAPSSVLDTVLLRPKEKNDVEYSETQELLRTEIVNPLRIYGVYCATKIMKLKRI 658
Db 181 TLFCLFAPSSVLDTVLLRPKEKNDVEYSETQELLRTEIVNPLRIYGVYCATKIMKLKRI 240
Qy 659 LEKVEAASGTSSEKDPPEFLNIFLPHLLRVEPLLKIRSAQKQVQDCYFYQIFMEKNEKV 718
Db 241 LEKVEAASGTSSEKDPPEFLNIFLPHLLRVEPLLKIRSAQKQVQDCYFYQIFMEKNEKV 300
Qy 719 GVPTIQQLLEWSFNSNLKFAEAPSCLIQMPRGKDFPKFKIPPSLELNTDLELTP 778
Db 301 GVPTIQQLLEWSFNSNLKFAEAPSCLIQMPRGKDFPKFKIPPSLELYIYILKTLTP 360
Qy 779 RQCRICGLAMYECCRECYDDPDIDISAGKIKOPCKTCNTQVHLHPKRLNHNKYNPVSPLKDL 838
Db 361 DSAGYVVEGLQCMSVENATTIRTSAGKIKOPCKTCNTQVHLHPKRLNHNKYNPVSPLKDL 420
Qy 839 DMDWRHGCIPQNNMELFAVLCIETSHYVAFVKYKDDSAWLFDFDSMADRDGGQNGFNIPQ 898
Db 421 DMDWRHGCIPQNNMELFAVLCIETSHYVAFVKYKDDSAWLFDFDSMADRDGGQNGFNIPQ 480
Qy 899 VTCPPEVGEVYKMSLEDLHSLDSRR 924
Db 481 VTCPPEVGEVYKMSLEDLHSLDSRR 506
```

RESULT 7

```
Q80VB3 PRELIMINARY; PRT; 313 AA.
ID Q80VB3 AC Q80VB3
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBDJ databases.
DR EMBL; BC049879; AAH49879.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 313 AA; 36198 MW; BED5662EAA7453DA CRC64;
```

Query Match 33.0%; Score 1662; DB 11; Length 313;
Best Local Similarity 97.8%; Pred. No. 1.4e-122;
Matches 306; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```
Qy 637 IVNPLRIYGVYCATKIMKLKILEKVEAASGFTSEKDPPEFLNIFHLLRVEPLLKIR 696
Db 1 IVNPLRIYGVYCATKIMKLKILEKVEAASGFTSEKDPPEFLNIFHLLRVEPLLKIR 60
Qy 697 SAQKQVQDCYFYQIFMEKNEKVGPVTIQQLLEWSFNSNLKFAEAPSCLIQMPRGKDF 756
Db 61 SAQKQVQDCYFYQIFMEKNEKVGPVTIQQLLEWSFNSNLKFAEAPSCLIQMPRGKDF 120
Qy 757 KLFKKIPPSLELNTDLELTPQCRICGGLAMYECCRECYDDPDIDISAGKIKOPCKTCNTQ 816
Db 121 KLFKKIPPSLELNTDLELTPQCRICGGLAMYECCRECYDDPDIDISAGKIKOPCKTCNTQ 180
Qy 817 VHLHPKRLNHNKYNPVSPLKDLPDMDWRHGCIPQNNMELFAVLCIETSHYVAFVKYKDDSD 876
Db 181 VHLHPKRLNHNKYNPVSPLKDLPDMDWRHGCIPQNNMELFAVLCIETSHYVAFVKYKDDSD 240
```



```
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RT Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RL "Annotation of Drosophila melanogaster genome.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RP [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnikier S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003628; AAN10738.1; -.
DR FlyBase; FBgn0032210; CG5603.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0004223; F:ubiquitin thioesterase activity; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH; 1.
SQ SEQUENCE 639 AA; 72128 MW; 6E29D3B09FF5E55B CRC64;

Query Match 22.6%; Score 1136.5; DB 5; Length 639;
Best Local Similarity 41.3%; Pred. No. 1.2e-80;
Matches 255; Conservative 103; Mismatches 202; Indels 57; Gaps 17;

QY 352 NGSSVDSQPSQSKNTWYDEVAEDPAKSLTEI-STDPDRSSPPLQPPVPVNSLTITENRFH 410
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 99
59 NHSAVDNDQ-----HLEDV-----DLADLGTNPKRPAGPAMILNNKSKTD----- 99

QY 411 SLPPSLTKMPTNGTNGIGHSPLSAQSVMEELNTAPVQESPPLAMPNGSHGLEVGSLAE 470
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 148
100 -----PNSNVDLILKASPILKIEPEELRTIADYQPLETPTGTE--LAIGSLVE 148

QY 471 VKENP-----PFGVIRWICQPG-LNEVLAGELEDEEC-----AGCTDGTFRTRFTTCAL 521
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 207
149 V-SNPGVCEDLYGVVRWIGIPPGQKQNLVGLGIEVEDSNLKNVWASDGRHNGVRLFTCHD 207

QY 522 KKAFLVKLSCRPDRFASLQP--VSNQIERCNSLARGCYLSEVVEENTPKM--EKEGL 577
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 267
208 GRAIFVFNRCADRRFADVDNISANRVSSNHAKKFGVADCPAIVGSIPIQIHNSDEL 267

QY 578 ETMIGKKGIQGHNSCYLDSTLFCILFAFSSVLDTVLLRPKEKNVVEYSETQELLRTI 637
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 327
268 ASICGKFKGIQGHNSCYLDALFMSFTFTSVFDSILYRRPGQDIRNYSVQKVLRLDI 327

QY 638 VNPLRIYGVYCATKIMKRLKILEKVAASGFTSEKDPBEFLNLPFHILRVEPLKIRS 697
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 387
328 VNPLRKNVFRDRVMKRLLEDQLSSVSLGTCCEKDPBEFLNLSQIMRVEPFLKLS 387

QY 698 AQQKVQDCFYQIFMEKNEKGVPTTQQLLEWSFINSNLKFAEAPSLIIQMPRFGKDFK 757
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 443
388 G-----QDSYFYQLFVEKDEKLTLPYQQQFEQSFHSSDILKLEVPSCFTIQMPRFGKNYK 443

QY 758 LFKKIPPSLELNTDLLETPQCRICGGLVMEYECREYDDPDISAG-KIKOFCKTCNTQ 816
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 503
444 MTPRILPSQVLDVTDIENSPPQSCUGKLAIEYECRDFGSLQAGSELCTAFCPKCLKT 503

QY 817 VHLHPLKRLNHNKVPNSLPKLDPLDWRHGCIFQNNMELFAVLICETSHYVAFYKYGK-DD 875
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 562
504 FHSIIKRTNHSVKIISPKFEKIMA-EHVMVPRLYMELFAVVICETSHYVAFVKSQSGPD 562

QY 876 SAWLPDSDMADRGGONGNINQVPTPCPEVGEYK-----MSLEDLHLSRSRIQGCARL 931
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 621
563 APWCFFDSMADRKGQNGYNIPEITCVBELTQWLSEBAGARSINET-STNDKVLPEHAKRI 621

QY 932 LCDAYMCMYQSPMSLY 948
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 638
622 FCDAYMCLYQSTDIMY 638

RESULT 11
Q81PCA PRELIMINARY; PRT; 550 AA.
ID Q81PCA
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 18, 2004, 15:16:48 ; Search time 8073 Seconds
(without alignments)
5095.075 Million cell updates/sec

Title: US-09-671-687A-3
Perfect score: 5034
Sequence: 1 MSGSLWQBKVTSPYWEERI.....RLLCDAYMCMYQSPMTSLYK 949

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DRV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09671687/runat_17092004_103034_2781/app.query.fasta_1.1095
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTWTF=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09671687@cgn_1_15932@runat_17092004_103034_2781 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4983	99.0	3540	9	BC012342 Homo sapi
2	4983	99.0	5414	9	AB020656 Homo sapi
3	4971.5	98.8	5371	9	AJ250014 Homo sapi
4	4942	98.2	4527	6	BD231207 Human cyt
5	4718.5	93.7	4314	10	AK122389 Mus muscu
6	4718.5	93.7	4501	10	BC042438 Mus muscu
7	4601	91.4	2845	6	AX883937 Sequence
8	4601	91.4	2845	6	BD160617 Primer fo
9	4601	91.4	2845	6	AK024348 Homo sapi
10	4600.5	91.4	241990	2	AC123449 Rattus no
11	4600.5	91.4	251132	2	AC098162 Rattus no
12	3676	73.0	2523	6	AR338799 Sequence
13	3433.5	68.2	2341	9	AK000187 Homo sapi
14	3022.5	60.0	2569	6	AX883681 Sequence
15	3022.5	60.0	2569	6	BD160470 Primer fo
16	3022.5	60.0	2569	9	AK024212 Homo sapi
17	2930	58.2	1954	9	AK056226 Homo sapi
18	2493	49.5	241630	2	AC131881 Rattus no
19	2468	49.0	2116	6	AK017216 Sequence
20	2468	49.0	2116	6	BD135433 Receptor
21	2108.5	41.9	215717	2	AC133651 Mus muscu
22	1736	34.5	251957	2	AC126867 Rattus no
23	1662	33.0	2426	10	BC049879 Mus muscu
24	1442.5	28.7	212531	2	AC145178 Gallus ga
25	1430.5	28.4	129025	2	AC145180 Gallus ga
26	1262	25.1	837	6	AX869679 Sequence
27	1262	25.1	837	6	BD149741 Primer fo
28	1215	24.1	758	6	AX867014 Sequence
29	1215	24.1	758	6	BD147076 Primer fo
30	1132.5	22.5	2068	3	AY071592 Drosophil
31	1111	22.1	151567	2	AC145503 Canis fam
32	1111	22.1	199277	2	AC145445 Canis fam
33	1109.5	22.0	208061	2	AC145254 Bos tauru
34	1108.5	22.0	168271	9	AC007728 Homo sapi
35	1097.5	21.8	141663	2	AC145018 Felis cat
36	1095.5	21.8	163319	9	HSA303140 Homo sapi
37	1085	21.6	155857	2	AC145174 Pan trogl
38	1085	21.6	178790	2	AC145238 Pan trogl
39	1078	21.4	194869	2	AC145003 Papio anu
40	1077.5	21.4	188742	2	AC138025 Mus muscu
41	1067	21.2	232509	2	AC134093 Rattus no
42	936	18.6	84367	3	AC005454 Drosophil
43	936	18.6	135011	2	AC020183 Drosophil
44	936	18.6	153505	3	AC009849 Drosophil
45	936	18.6	154779	3	AC092237 Drosophil

ALIGNMENTS

BC012342	3540 bp	mRNA	linear	PRI 04-OCT-2003
LOCUS	Homo sapiens cylindromatosis (turban tumor syndrome), mRNA (cDNA			
DEFINITION	clone MGC:19923 IMAGE:4552767), complete cds.			
ACCESSION	BC012342			
VERSION	BC012342.1 GI:15214433			
KEYWORDS	MGC.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.			
	1 (bases 1 to 3540)			
	Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,			
	Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,			
	Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,			
	Hopkins, R.F., Jordan, H., Moore, J., Max, S.I., Wang, J., Hsieh, F.,			
	Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,			
	Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,			
	Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,			
	Carninci, P., Prange, C., Raha, S.S., Lequellano, N.A., Peters, G.J.,			
	Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,			
	McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,			
	Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,			
	Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,			
	Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S.,			
	Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,			
	Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,			
	Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,			
	Butterfield, A.S., Krzyzanski, M.I., Skalska, U., Smalhus, D.E.,			
	Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.			
TITLE	Generation and initial analysis of more than 15,000 full-length			
	human and mouse cDNA sequences			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)			
MEDLINE	22388257			
PUBMED	12477932			
REFERENCE	2 (bases 1 to 3540)			
AUTHORS	Straussberg, R.			
TITLE	Direct Submission			
JOURNAL	Submitted (15-AUG-2001) National Institutes of Health, Mammalian			
	Gene Collection (MGC), Cancer Genomics Office, National Cancer			
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,			
	USA			
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov			
COMMENT	Contact: MGC help desk			
	Email: cgabbs-remail.nih.gov			
	Tissue Procurement: ATCC			
	cDNA Library Preparation: Rubin Laboratory			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: Genome Sequence Centre,			
	BC Cancer Agency, Vancouver, BC, Canada			
	info@bcgsc.bc.ca			
	Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,			
	Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,			
	Leticia Hsiao, Martin Krzyzanski, Reta Kutache, Oliver Lee, Soo			
	Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven			
	Ness, Fawan Pandolf, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline			
	Schein, Duane Smalley, Michael Smith, Lorraine Spence, Jeff Stott,			
	Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,			
	George Yang, Scott Zuyderduyn, Marco Marra.			
FEATURES	Clone distribution: MGC clone distribution information can be found			
source	through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov			
	Series: IRAL Plate: 29 Row: c Column: 5			
	This clone was selected for full length sequencing because it			
	passed the following selection criteria: matched mRNA gi: 14165257.			
	Location/Qualifiers			
	1..3540			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="MGC:19923 IMAGE:4552767"			
	/tissue_type="Uterus, leiomyosarcoma"			
	/clone_lib="NIH_MGC_46"			
	/lab_host="DH10B-R"			
	/note="Vector: pOTB7"			
	1..3540			
	/gene="CYLD"			
	/note="synonyms: CDMT, CYLD1, EAC, KIAA0849, HSPC057"			
	/db_xref="LocusID:1540"			
	/db_xref="MIM:605018"			
	277..3138			
	/codon_start=1			
	/product="CYLD protein"			
	/protein_id="AAH12342.1"			
	/db_xref="GI:15214434"			
	/db_xref="LocusID:1540"			
	/translation="MSSGLMQEKNVTSYVWRIIFVLLLOEQSVTDKQTKLLKVPKG			
	SIGYIQDRSVGHSRIIPSAKGGKKNQIGIDVGPVKVQLRSGEEKFPVGRFRGLAERTVSG			
	IPFGVLEBEGRGQFTDVYQKQLQDCEDCVFVADLKLELLEDDDDTALESYAG			
	PQDMQVLELPLEINRSVLKGETIESGTIVFCVLPKESLGIFVGVMDNPIGNW			
	KGRFDVQLCSFACVETILLINDIIPESVTOERRPKLAFMSRGVGGKGSNNKP			
	KATGSTDPGNRSSELFVTLNGSSVDSPQSKNTWYIDEVARDPAKSLTEISDF			
	DRSSPLPPPPVNSLTTRPHSLFSLTKMNTNGSIGHSPLSLSAQSVMBELNTP			
	VOESPPLAMPNGSHGLEVSLAEVKEKNPPFYGVIRWIGQPEGLNEVLAGELESCA			
	GCTDGTFRGTYFTCALKALFVLKLSRCPDRSFASLQVSNQIERCNSLAFGGYLS			
	VVEENTPPMEKEGLEIMIGKKGIQGHYNSCYLSDTLFCLFAFSSVLDVLLRPKEK			
	NDVEYSTQELLRIEIVNPLRIYGVCACTIKMLRKILEKVEAASGFTSEKDPREF			
	LNILPHILRVPELLKIPSAQKQVDCYFOIFMEKNEKVGVTTCQLLEWFSINSL			
	KFAEASCLIIQMPREGDKFLFKIIPFSLNITDLEDTPRQRCICGLAYECRE			
	CYDDPDISAGIKQCKTCTQVHLHPKLNHNKYNPVSPLKDLPMWDMRHGCI PQQNM			
	ELFVLCIETSHYVAFVKYKDDSAWLFDSMADRGQGNFNIQVTPCPPEVGEYLK			
	MSLEDLHSDLSRRIIQGCARRLLCDAYMCWQSPMTSLVK"			
	745..867			
	/note="CAP_GLY; Region: CAP-Gly domain. CAP stands for			
	cytoskeleton-associated proteins"			
	/db_xref="CDD:pfam01302"			
	1720..2025			
	/note="NIP100; Region: COG5244, NIP100, Dynactin complex			
	subunit involved in mitotic spindle partitioning in			
	anaphase B [Cell division and chromosome partitioning]"			
	/db_xref="CDD:COG5244"			
ORIGIN	US-09-671-687A-3 (1-949) x BC012342 (1-3540)			
Alignment Scores:				
Pred. No.:	0	Length:	3540	
Score:	4983.00	Matches:	948	
Percent Similarity:	99.48%	Conservative:	0	
Best Local Similarity:	99.48%	Mismatches:	1	
Query Match:	98.99%	Indels:	4	
DB:	9	Gaps:	4	
QY	1	MetSerSerGlyLeuTrpSerGlnGluysValThrSerProTyrTrpGluArgile	20	
Db	277	ATGAGTTTCAGGCTTATGGAGCCCAAGAAAAGTCACTTCACTTGGGAAGCGGATT	336	
QY	21	PheTyrLeuLeuGlnGluCysSerValThrAspLysGlnThrGlnLysLeuLys	40	
Db	337	TTTTACTTCTTCTTCAAGATGCAGCGTTACAGACAAACAAACAAAGCTCCTTAA	396	
QY	41	ValProLysGlySerIleGlyGlnTyrIleGlnAspArgSerValGlyHisSerArgile	60	
Db	397	GTACCGAAGGGAAGTATAGACAGATATATCAAGATCGTTCGTGGGGCATTCAGATT	456	
QY	61	ProSerAlaLysGlyLysLysAsnGlnIleGlyLeuLysIleLeuGluGlnProHisAla	80	
Db	457	CCTTCTGCAAAAGGCAAGAAAATCAGATTGATTAATAATTTCTAGAGCAACCTCATGA	516	
QY	81	ValLeuPheValAspGlu---AspValValGluIleAsnGluLysPheThrGluLeuLeu	99	
Db	517	GTCTCTTTTGTGATCAAAAGGATGTTGTAGAGATAAATAAGTTCACAGATTACTT	576	
QY	100	LeuAlaIleThrAsnCysGluGluArgPheSerLeuPheLysAsnArgAnArgLeuSer	119	

577	TTGCAATATCCAAATTGTGAGGAGAGGTTACAGCTGTTTAAAAACAGAAACAGACTAAGT	636
120	LysGlyLeuGlnIleAspValGlyCysProValLysValGlnLeuAArgSerGlyGluGlu	139
637	AAAGGCCTCCAAATAGACGTGGGCTGTCTGTGAAAGTACAGCTGAGATCTGGGGAAGAA	696
140	LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluAArgThrValSerGly	159
697	AAATTTCTCGAGTTGTACGCTTCAGAGGACCCCTGTTAGCAGAGAGACAGACTCTCCGGA	756
160	IlePhePheGlyValGluLeuGluGluGluValArgGlyGlnGlyPheThrAspGlyVal	179
757	ATATCTTTGGAGTTGAATTCTCTGGAAGAAGGTCGTGCTCAAGGTTTCACTGACGGGGTG	816
180	TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAsp	198
817	TACCAAGGGAACACAGCTTTTTCAGTGTGATGAAGATTGTGCGGTGTTGTTCATTGGAC	876
199	LysLeuGluLeuIleGluAAspAspThrAlaLeuGluSerAspTyrAlaGlyProGly	218
877	AAGCTAGAAGCTCATAGAAGATGATGACATGTCATTGGAAGTGTATTACGCAGGTCCTGGG	936
219	AspThrMetGlnValGluLeuProProLeuGluIleLeuSerArgValSerLeuLysGly	238
937	GACAAATGACAGTCTGAACCTCTCTCTTTGGAAATAAACTCCAGAGTTTCTTTGAAGGTT	996
239	GlyGluThrIleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSer	258
997	GGAAACAATAGAACTCTGGNACAGTTATATTCTGTGATGTTTTCAGGAAAGAAGAAGC	1056
259	LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTrpAspGlyArgPhe	278
1057	TTAGGATATTTTGTGTGTGTGGACATGGATAAACCTATTGGCAACTGGGATGGAAGATTT	1116
279	AspGlyVal---LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsn	297
1117	GATGGAGTGCAGCTTTTGTAGTTTTCGTGTGTGTGAAGAGTCAAAATCTATTTCACATCAAT	1176
298	AspileIleProGluSerValThrGlnGluArgArgProProLysLeuAlaPheMetSer	317
1177	GATATCATCCAGAGAGTGTGACGCGAGGAAGGAGGCTCCCAAACTTGCTTTTATGTCA	1236
318	ArgGlyValGlyAspLysGlySerSerHisAsnLysProLysAlaThrGlySerThr	337
1237	AGAGCTGTGGGCAAAAGGTTATCCAGTCATAATAAAACCAAGGCTCAGAGATCTACC	1296
338	SerAspProGlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsnGlySerSerVal	356
1297	TCAGACCTCGAATAAGAAAACAGATCTGAATATATTTTATACCTTAAATGGGTCTCTGTT	1356
357	AspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluValAlaGluAsp	376
1357	GACTCACACACACANTCCAATCAAAAATACATGGTACATTGATGAAGTTGCAGAGAC	1416
377	ProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSerProProLeuGln	396
1417	CCTGCAAAATCTCTTACAGAGATATCTACAGACTTTCACCGTCTCTCCACCACCTCCAG	1476
397	ProProProValAsnSerLeuThrThrGluAsnArgPheHisSerLeuProPheSerLeu	416
1477	CCTCCTCCTGTGAATCTCACTGACCAACGAGAACAGATTCCACTCTTTTACCATTTCAGTCTC	1536
417	ThrLysMetProAsnThrAsnGlySerIleGlyHisSerProLeuSerLeuSerAlaGln	436
1537	ACCAAGATGCCCAATACCAATGGGAAGTATTGGCCACAGTCCACTTTCTCTGTACGCCAG	1596
437	SerValMetGluGluLeuAsnThrAlaProValGlnGluSerProProLeuAlaMetPro	456
1597	TCTGTAATGGAAGACTAAACACTGTCACCCGCTCCAAGAGAGTCCACCTTGGCCATGCGCT	1656
457	ProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValLysGluAsnProPro	476
1657	CCTGGGAACTCATGTCTAGAACTGGGCTCATTTGGCTGAAGTTAAGGAGAACCCTTCCT	1716

Qy	477	PheTyrGlyValIleAleArgTfPrlIleGlyGlnProProGlyLeuAsnGluValLeuAlaGly	496
Db	1717	TTCTATGGGGTAATCCGTTGGATCGGTGACCCACACAGGACTGAATGAAGTGCCTCGCTGGA	1776
Qy	497	LeuGluLeuLeuAspGluCysAlaGlyCysThrAspGlyThrPheArgGlyThrArgTyr	516
Db	1777	CTGGAAC TGGAAGATGAGTGTGACGGCTGTACCGATGGAACCTTCAGAGGCATCCGGTAT	1836
Qy	517	PheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArgProAspSer	536
Db	1837	TTCCACTGTGCTCTGAAGAAGCGCTGTTTGTGAACCTGAAGAGCTGCAGCGCTGACTCT	1896
Qy	537	ArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsnSerLeuAlaPhe	556
Db	1897	AGGTTTGCATCATTCAGCGCGGTTTCCAACTCAGATTGAGCGCTGTAACTCTTTAGCATTT	1956
Qy	557	GlyGlyTyrLeuSerGluValValGluGluAsnThrProProLysMetGluLysGluGly	576
Db	1957	GGAGGCTACTTAAAGTGAAGTAGTAGAAGAAAACTCCCAAAAAATGGAAAAAAGAGGC	2016
Qy	577	LeuGluIleMetIleGlyLysLysLysGlyIleGlnGlyHisTyrAsnSerCysTyrLeu	596
Db	2017	TTGGAGATAATGATTGGGAAGAGAAGGCATCCAGGTCATTACAATCTCTGTACTTAA	2076
Qy	597	AspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThrValLeuLeuArg	616
Db	2077	GACTCAACCTTATTCCTGCTTATTTCGTTTTAGTTCTGTTCTGACACTGTGTTACTTTAGA	2136
Qy	617	ProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeuArgThrGlu	636
Db	2137	CCCAAGAAAGAACGATGTAGAAATATATAGTGAACCAACAGAGCTACTTGAGCAGACAA	2196
Qy	637	IleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrLysIleMetLysLeuArg	656
Db	2197	ATTGTTAATCCTCTGAGAATATATGATATGTGTGTGCCCAAAAAATATGAAACATGAGG	2256
Qy	657	LysIleLeuGluLysValGluAlaAlaSerGlyPheThrSerGluGluLysAspProGlu	676
Db	2257	AAAATACTTGAAAGGTGGAGGCTGCATCAGGATTTACCTCTGAAGAAAAAGATCCTGAG	2316
Qy	677	GluPheLeuAsnIleLeuPheHisHisIleLeuArgValGluProLeuLeuLysIleArg	696
Db	2317	GAATTTCTCAATATCTGTTTCATCATATATTTAAGGTTAGAACTTTGCTTAAAAAAAAGA	2376
Qy	697	SerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAsnGlu	716
Db	2377	TCAGCAGGTCAAAGGTCAAGAAGTGTATCTCTATCAAAATTTTATGGAAGAAAAATGAG	2436
Qy	717	LysValGlyValProThrIleGlnGlnLeuLeuGluTyrPheIleAsnSerAsnLeu	736
Db	2437	AAAGTTGGCGTTCCCAATTCAGCAGTTGTAGAAATGCTCTTTTATCAACAGTAACTTG	2496
Qy	737	LysPheAlaGluAlaProSerCysLeuIleIleGlnMetProArgPheGlyLysAspPhe	756
Db	2497	AAATTTTGCAGAGCCACATCATCTCTGATTATTCAGATGCTCGAATTTTGAAAAAGACTTT	2556
Qy	757	LysLeuPheLysLysIlePheProSerLeuGluLeuAsnIleThrAspLeuLeuGluAsp	776
Db	2557	AAACTATTTAAAAAAAATTTTTCTCTCTGGAATTAATAATATAACAGATTTTACTTTGAAGAC	2616
Qy	777	ThrProArgGlnCysArgIleCysGlyGlyLeuAlaMetTyrGluCysArgGluCysTyr	796
Db	2617	ACTCCAGACAGTCCGGATATGTGGAGGCTTGCATGTATGAGTGTAGAGAAATGCTTAC	2676
Qy	797	AspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAsnThrGln	816
Db	2677	GACGATCCGGACATCTCAGCTGGAAGAAATCAAGCAGTTTGTGAAAACTTGCACACTCAA	2736
Qy	817	ValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeuProLysAsp	836
Db	2737	GTCCACTTCATCCGAAGAGGCTGAATCATATAATATAACCCAGTGTCACTTCCCAAGAC	2796


```

QY 837 LeuProAspTTPAspTTPArgHisGlyCysIleProCysGlnAsnMetGluLeuPheAla 856
Db 2797 TTAACCGAGTGGAGACACGGCTGCATCCCTTGCAGAAATATGAGTTATTGCT 2856
QY 857 ValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspSer 876
Db 2857 GTTCTCTGATAGAAACAGCCACTATGTTGCTTTTGTGAAGTATGGAAGACGATTCT 2916
QY 877 AlaTTPLeuPhePheAspSerMetAlaAspArgAspGlyGlyGlnAsnGlyPheAsnIle 896
Db 2917 GCCTGGCTCTTTTACAGACATGCGGATGCGGATGGTGGTGCAGAAATGGCTTCAACATT 2976
QY 897 ProGlnValThrProCysProGluValGlyGluTyrLeuLysMetSerLeuGluAspLeu 916
Db 2977 CTTCAAGTCAACCCATGCCAGAGTAGAGAGTACTTGAAGATGCTCTGGAAGACCTG 3036
QY 917 HisSerLeuAspSerArgArgIleGlnGlyCysAlaArgArgLeuLeuCysAspAlaTyr 936
Db 3037 CATTCCTGGACTCAGGAGATCCAAAGCTGTGCAGGAAGACTGCTTTGTGATGCATAT 3096
QY 937 MetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949
Db 3097 ATGTGCATGTACAGATCCAAATGAGTTTGTACAAA 3135

RESULT-2
LOCUS AB020656 5414 bp mRNA linear PRI 17-MAY-2001
DEFINITION Homo sapiens mRNA for KIAA0849 protein, partial cds.
ACCESSION AB020656
VERSION AB020656.2 GI:14133220
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 5414)
AUTHORS Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hiroseawa,M., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
JOURNAL DNA Res. 5 (6), 355-364 (1998)
MEDLINE 99156230
PUBMED 10048495
REFERENCE 2 (bases 1 to 5414)
AUTHORS Ohara,O., Suyama,M., Kikuno,R., Nagase,T. and Ishikawa,K.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1998) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnaiffo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)
COMMENT On May 17, 2001 this sequence version replaced gi:4240186.
FEATURES
source
1..5414
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="fh04363"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
/dev_stage="adult"
/note="vector:pBluescriptII SK plus. This sequence is replaced that of hk05904 cDNA as a representative cDNA sequence for KIAA0849"
gene
1..5414
/gene="KIAA0849"
<426..3308
/gene="KIAA0849"
/codon_start=1
/product="KIAA0849 protein"
/protein_id="BAA74872.2"
/db_xref="GI:14133221"

```

```

/translaton="YFENITMSSGLWSQEKVTSYWBBERIFYLLIQECSVTDKOTQK
LLKVPKSGIYQIDORSVGHRSIPSAKGNQIGLILKLEOPHAVLFDVDEKQVVEINEK
FTELLATLNGEERESLKNRNLKGLQIDQCPVKVQLRSGEKFPQVWVRPGLL
AERTVSGIFFGVELLESEGQGTGCVTQKQLFCQDEDCGVFVLDKLELLEDDDTA
LESDYAGPDIMQVELPLEINRSVSLKVGTEISGTVIFCDVLPGKESLGSYFVGDM
DNPIGNWDRFDGVLQCSFACVESTILLIINDIIPESTQERRPPKLAFPMMSGDGK
SSHNKPKATQSDPNNRSELFTLNGSVSDQSQSKNTWYIDVADDPKSL
TEISTDFRSPPPLQPPVNSLTENRFLSLPFLTKMPTKNTSGHSLSLSAQSV
BELTAPQESPPPLAMPNGSHGLEVAEKENPFPFVIRWIRGQPLNEVLGL
ELEDCACCTDGTFRGTFTYFCALKKALFKLSCRPDRFASLQPVNSQIERCNLA
FGGLYSEVEVENTPPMEKGLIEMIGKKGIQGHNSCYLSTLLCFASFSSVLDTV
LLRPEKNDEYIYSQTELLRTIYNPLRIYGVCAIKMLKILKEVTAASGFTSE
EKDPEFINILFPHILRVEPLLKIRSGAQVQDQCYFOIFMEKNEKVGVTYQQLLEW
SFINSNLFAEAPASCLIIQMPRFKQDKLKKIIPSLLEINIITDLEDTQPCRICGL
ANYECRECYDDPDISAGIKOPCKTCNTQVHLHPKRLNHNPNVSLPKDLPDWDRHG
CIPQNMELFACIETSHYVAFVYKDDSAWLFDFSDNADRGGONGFNIFQVTPCP
EVGEYLMKSLDLHSLDSRRITQGCARRILLCDAYCMYQSPMSLYK"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 5414
Score: 4983.00 Matches: 948
Percent Similarity: 99.48% Conservative: 0
Best Local Similarity: 99.48% Mismatches: 1
Query Match: 98.99% Indels: 4
DB: 9 Gaps: 4
US-09-671-687A-3 (1-949) x AB020656 (1-5414)
QY 1 MetSerSerGlyLeuTTPSerGlnGlyValThrSerProTyrTTPGluGluArgIle 20
Db 447 ATGAGTTTCAAGCTTATGAGCCCAAGAAAAGTCACTTCACTCCCTACTGGGAAGCGGATT 506
QY 21 PheTyrLeuLeuGlnGluCysSerValThrAspLysGlnThrGlnLysLeuLys 40
Db 507 TTTTACTTCTTCTTCAAGAATGCAGCGTTACAGACAAACAAACAAAGCTCTCTAAA 566
QY 41 ValProLysGlySerIleGlyGlnTyrIleGlnAspArgSerValGlyHisSerArgIle 60
Db 567 GTACCGAAGGGAAGTATAGGACAGTATATTCAGATCGTCTGTGGGGCATTTCAAGGATT 626
QY 61 ProSerAlaLysGlyLysLysAsnGlnIleGlyLeuLysIleLeuGluGlnProHisAla 80
Db 627 CCTTCTGCAAAAGCGCAAGAAAATCAGATTGATTAATAATTTCTAGAGCAACCTCATGCA 686
QY 81 ValLeuPheValAspGlu---AspValValGluIleAsnGluLysPheThrGluLeuLeu 99
Db 687 GTTCTCTTTGTTGATGAAAAGGATGTTGTAGAGATAAATGAAAAGTTTCAAGAGTTACTT 746
QY 100 LeuAlaIleThrAsnGlyGluArgPheSerLeuPheLysAsnArgAsnArgLeuSer 119
Db 747 TTGGCAATTTACCAATTTGTGAGGAGAGGTTTCAAGCTCTTTTAAACACAGAACACTAAGT 806
QY 120 LysGlyLeuGlnIleAspValGlyCysProValLysValGlnLeuArgSerGlyGluGlu 139
Db 807 AAAGGCTTCCAATAGACGTGGGCTGTCTCTGTAAGTACAGCTGAGATCTGGGGAGAA 866
QY 140 LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGly 159
Db 867 AAATTTCTTGGAGTTGTACGCTTACAGAGACCCCTGTTAGCAGAGAGGACAGCTCTCCGA 926
QY 160 IlePhePheGlyValGluLeuLeuGluGlyArgGlyGlnGlyPheThrAspGlyVal 179
Db 927 ATATTTCTTTGGAGTTGAATTTGTGGAAGAGGTCGTGGTCAAGGTTTCTACTGACGGGGTG 986
QY 180 TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAsp 198
Db 987 TACCAGGGAACACGCTTTTTCAGTGTGATGAAGATTGTGGCTGTTTGTGATTGGAC 1046
QY 199 LysLeuGluLeuIleGluAspAspThrAlaLeuGluSerAspTyrAlaGlyProGly 218
Db 1047 AAGCTAGAACTCATAGAACATGATGACACTGCATTCGAAAGTATTACGACGAGTCTCTGG 1106

```

QY 219 AspThrMetGlnValGluLeuProProLeuGluIleAsnSerArgValSerLeuLysGly 238
DB 1107 GACAAATGCGAGTCGAACTTCCTCTTGGAAATAAACTCCAGAGTTCTCTTTGAAGGTT 1166
QY 239 GlyGluThrIleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSer 258
DB 1167 GGAGAAACAATAGNATCTGGACAGTTATATCTGTGATGTTTTGCCAGGAAAGAAGC 1236
QY 259 LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTrpAspGlyArgPhe 278
DB 1227 TTAGGATATTTGTGTGTGGACATGGATAACCTATTGGCAACTGGGATGGAGATT 1286
QY 279 AspGlyVal---LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsn 297
DB 1287 GATGGAGTGCAGCTTTGTAGTTTTGCGTGTGTGAAAGTACAATTTCTATTGACATCAAT 1346
QY 298 AspileileProGluSerValThrGlnGluArgArgProProLysLeuAlaPheMetSer 317
DB 1347 GATATCATCCAGAGAGTGTACCCAGAAAGGAGGCTCCCAAACTTGCCTTTATGTCA 1406
QY 318 ArgGlyValGlyAspLysGlySerSerHisAsnLysProLysAlaThrGlySerThr 337
DB 1407 AGAGGTGTTCGGGACAAAGGTTTCATCCAGTCATAATAAACCAAGGCTACAGGATCTACC 1466
QY 338 SerAspProGlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsnGlySerSerVal 356
DB 1467 TCAGACCCTGGAATAGAAACAGACTCTGAATTTATTTATACCTTAAATGGGCTCTCTGTT 1526
QY 357 AspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluValAlaGluAsp 376
DB 1527 GACTCACAAACCAATCCAAATCAAAATAACATGTTGACATTTGATGAAGTTGCAGAAAGC 1586
QY 377 ProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSerProProLeuGln 396
DB 1587 CCTGCAAAATCTCTACAGAGATATCTACAGACTTTGACCGTCTTCACCACTCCAG 1646
QY 397 ProProProValAsnSerLeuThrThrGluAsnArgPheHisSerLeuProPheSerLeu 416
DB 1647 CCTCCTCCTGTGAACCTCACTGACCACCGAGAACAGATTCCACTCTTTTACCATTCACTCTC 1706
QY 417 ThrLysMetProAsnThrAsnGlySerIleGlyHisSerProLeuSerLeuSerAlaGln 436
DB 1707 ACCAAGATGCCAATAACCAATGGAAGTATTGGCCACAGTCCACTTCTCTGTGAGCCAG 1766
QY 437 SerValMetGluGluLeuAsnThrAlaProValGlnGluSerProProLeuAlaMetPro 456
DB 1767 TCTGTAATGGAAGAGCTAAACACTGCACCCGTCAGAGAGTCCACCTTTGGCCATGCCT 1826
QY 457 ProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValLysGluAsnProPro 476
DB 1827 CCTGGGAACTCATGGTCTAGAAAGTGGGCTCATTTGGCTGAAGTTAAGGAGAACCTCTCT 1886
QY 477 PheTyrGlyValIleArgTrpIleGlyGlnProProGlyLeuAsnGluValLeuAlaGly 496
DB 1887 TTCTATGGGTAATCCGTTGGATCGGTGAGCCAGCCAGGACTGAATGAAGTCTCGCTGGA 1946
QY 497 LeuGluLeuLysGluCysAlaGlyCysThrAspGlyThrPheArgGlyThrArgTyr 516
DB 1947 CTGGAATCGGAAGATGAGTGTGAGGCTGTACCGATGGAACTTCAGAGGACTTCGGTAT 2006
QY 517 PheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArgProAspSer 536
DB 2007 TTCACCTGTGCTTGAAGAGGCGCTGTTGTGAACTGAAGAGCTGCAGGCTTACTCT 2066
QY 537 ArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsnSerLeuAlaPhe 556
DB 2067 AGCTTTGCATCAATTGCAGCGGTTTCCAATCAGATTGAGCGCTGTAATCTCTTAGCATTT 2126
QY 557 GlyGlyTyrLeuSerGluValValGluGluAsnThrProProLysMetGluLysGluGly 576
DB 2127 GGAGGCTACTTAAGTAGTAGAGAAATACTCCACCAAAATGGAAGGAGGAGG 2186
QY 577 LeuGluIleMetIleGlyLysLysGlyIleGlnGlyHisTyrAsnSerCysTyrLeu 596

DB 2187 TTGGAGATAATGATTTGGGAAGAAAGGCAATCCAGGTCATTACAAATTTCTTGTACTTTA 2246
QY 597 AspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThrValLeuLeuArg 616
DB 2247 GACTCAACCTTATTTCTGCTTATTTGCTTTTAGTTCTGTCTGGACACTGTGTACTTAGA 2306
QY 617 ProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeuArgThrGlu 636
DB 2307 CCCAAAGAAAGAACGATGTAGATAATTTATAGTGAACCCCAAGAGCTACTGAGGACAGAA 2366
QY 637 IleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrLysIleMetLysLeuArg 656
DB 2367 ATTGTTAATCTCTCAGAAATATATGATATGTGTGCGCACAAAATTTATGAACTGAGG 2426
QY 657 LysIleLeuGluLysValGluAlaSerGlyPheThrSerGluGluLysAspProGlu 676
DB 2427 AAAATACTTGAAGAGGTGGAGGCTGCATCAGATTTCCTCTGAAAGAAAAGATCCTGAG 2486
QY 677 GluPheLeuAsnIleLeuPheHisIleLeuArgValGluProLeuLeuLysIleArg 696
DB 2487 GAATTTCTGAATATTTCTGTTTCATCATATTTTAAAGGTAGAACCTTTGCTAAAAATAGA 2546
QY 697 SerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAsnGlu 716
DB 2547 TCAGCAGGTCAAAAGGTACAGATTGTTACTTCTATCAAAATTTTATGCAAAAAAATGAG 2606
QY 717 LysValGlyValProThrIleGlnGlnLeuLeuGluTrpSerPheIleAsnSerLeu 736
DB 2607 AAAGTTGGCGTTCACCAATTCAGCAGTTGTAGAAATGTTCTTTTATCAACAGTAACCTG 2666
QY 737 LysPheAlaGluAlaProSerCysLeuIleGlnMetProArgPheGlyLysAspPhe 756
DB 2667 AAATTTCAGAGGACCATCATGTCTGATTAATTAGATGCCTCGATTTGGAAAAGACTTT 2726
QY 757 LysLeuPheLysLysIlePheProSerLeuGluLeuAsnIleThrAspLeuLeuGluAsp 776
DB 2727 AAACATAATTAATAATTTTCTCTCTGAAATTAATAATAACAGATTTACTTGAAGAC 2786
QY 777 ThrProArgGlnCysArgIleCysGlyGlyLeuAlaMetTyrGluCysArgGluCysTyr 796
DB 2787 ACTCCAGACAGTCCCGGATATGTGGAGGGCTTGCAATGTATGAGTGTAGAGAATGCTAC 2846
QY 797 AspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAsnThrGln 816
DB 2847 GACATCCGACATCTCAGCTGGAATAATCAAGCAGTTTGTAAACCTGCAACACTCAA 2906
QY 817 ValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeuProLysAsp 836
DB 2907 GTCCACCTTCATCCGAAGAGGCTGAATCATATAATAATAAACCCAGTGTCACTTCCCAAGAC 2966
QY 837 LeuProAspTrpAspTrpArgHisGlyCysIleProCysGlnAsnMetGluLeuPheAla 856
DB 2967 TTACCCGACTGGGACTGGAGACACGGCTGCATCCTTGGCCAGAAATATGGAGTTATTGCT 3026
QY 857 ValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspSer 876
DB 3027 GTTCTCTGCATAGAAACAGCCACTATGTGCTTTTGTGAAGTATGGGAAGGACGATCT 3086
QY 877 AlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlyGlnAsnGlyPheAsnIle 896
DB 3087 GCCTGGCTCTCTTTGACAGCATGGCCGATCGGATGGTGTGTCAGATGGCTTCACACTT 3146
QY 897 ProGlnValThrProCysProGluValGlyGluTyrLeuLysMetSerLeuGluAspLeu 916
DB 3147 CCTCAAGTCAACCCATGCCCAGAAAGTAGGAGTACTTGAAGATGTCTCTGGAAGACCTG 3206
QY 917 HisSerLeuAspSerArgArgIleGlnGlyCysAlaArgArgLeuLeuCysAspAlaTyr 936
DB 3207 CATTTCTTGGACTCCAGGAGATCCAAAGGCTGTGCAGGAAGACTGCTTTGTGTCATAT 3266
QY 937 MetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949

Db	3267	ATGTGATGACAGAGTCCACAAATGAGTTGTACAAA	3305
RESULT 3			
LOCUS	HS250014	5371 bp mRNA linear	PRI 16-JUN-2000
DEFINITION	Homo sapiens mRNA for Familial Cyndromatosis cyld gene.		
ACCESSION	AJ250014		
VERSION	AJ250014.1	GI:8250235	
KEYWORDS	cyld gene; Familial Cyndromatosis.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 Bignell,G.R., Brown,C., Biggs,P.J., Lakhani,S.R., Jones,C., Hansen,J., Blair,E., Hofmann,B., Siebert,R., Turner,G., Evans,D.G., Schrandt-Stumpel,C., Beemer,F.A., Van Den Ouweland,A., Halley,D., Delpech,B., Cleveland,M.G., Leigh,I., Leisti,J., Rasmussen,S., Wallace,M.R., Fenske,C., Banerjee,P., Oiso,N., Chaggar,R., Merritt,S., Leonard,N., Huber,M., Hohl,D., Chapman,P., Burn,J., Swift,S., Smith,A., Ashworth,A. and Stratton,M.R.		
TITLE	Identification of the familial cyndromatosis tumour-suppressor gene		
JOURNAL	Nat. Genet. 25 (2), 160-165 (2000)		
MEDLINE	20296617		
PUBMED	10835629		
REFERENCE	2 (bases 1 to 5371)		
AUTHORS	Stratton,M.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-NOV-1999) Stratton M.R., Cancer Genetics, Institute of Cancer Research, 15 Cotswold Rd, Sutton,, Surrey. SM2 5NG., UNITED KINGDOM		
FEATURES	Location/Qualifiers		
source	1..5371		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map="16q12 - q13"		
gene	1..5371		
	/gene="cyld"		
CDS	392..3262		
	/gene="cyld"		
	/function="tumour suppressor/recessive oncogene"		
	/note="Familial Cyndromatosis Gene"		
	/codon_start=1		
	/evidence=experimental		
	/protein_id="CAB93533.1"		
	/db_xref="GI:8250236"		
	/db_xref="GOA:O9NOC7"		
	/db_xref="SPTREMBL:O9NOC7"		
	/translation="MSSGLWSQEKVTSFYWEERFYLLQBCSVTDKQTKLLKVPKG SICYQIDRSVGHRSIPSAKGNKQIGLILKQPHAVLFDVDEKDVINEKFTLELLIA ITNCEERFLPNRLRSLKGLQIDVCPVKQLRSGEKFPGVVFRFGPLLAERTVSG IFFGVELLEEGRGQGTGQVYKQGLFQDCDECGVFVALDKLEIEDDDTALERSDYG PGDMOVELPPEEINRSVLKVGTEISGTVIFCDVLPGKESLCYFVGVDMNPINW DGRFDGVLQCFACVSESTILLHNDIIPALSESVTQERPPKLAFLMRSGVDCXSSSH NPKATGTSIDSGNRSSELFTLNGSSVDSQPSQKNTWYIDEVARDPAKSLTEIS TDFRSPLTPPPVNSLTTFNHSPLSPKNTNGTSIGHSPLSLSAQSVNEELN TAPVQSPPLAMPNGSHLEVGSLAEYKPNPPFVIRWIGQPPGLNEVLAGELED ECAGTDTGFRTRYFTCALKALFVKLKSPPRFSALQPSVSNQIERCNSLAFGY LSEVNDPPPMKEGELMIGKKKGIGHYNSCYLDSTFLCLFAPFSVLDVTLRLP KEKNDVEYSETEQLLRTEIYNPLRIYGVCACTIMKLKILEKVEAASGFTSEKDP EPLNLTFHILRLVPLLKIRSAQKVQDCFYQIFMEKNEKVGVPFIQQLLEWSPFN SNLKFAEASCLIIQMPRAFQDKFLFKIFPSLEINITDLEEDTPQRCRCGGIAMEY CRECYDDPDIISAGKIQFCKTQNTQHLHKLRLNHNPNVSLPKDLPDWRHGCTPC QNMELFVLVCIETSHYAFVKYKDDSAWLFDFDSWADRDGGQNGFNIPQVTPCPVEGE YLMKSLDLHSLDSRRIQGCARRLLCDAYMCWYQSPTMSLYK"		
ORIGIN			
Alignment Scores:			
Pred. No.:	0	Length:	5371
Score:	4971.50	Matches:	948

Db 1352 TTATGTCAAGAGGTGTGGGACAAAGGTTTCATCCAGTCATATAATAACCAAGGCTACA 1411
Qy 335 GlySerThrSerAspProGlyAsnArg- --ArgSerGluLeuPheTyrThrLeuAsnGly 353
Db 1412 GGATCTACCTCAGACCCCTGGAATAGAAACAGATCTGAATATATTTATATACCTTAATGGG 1471
Qy 354 SerSerValAspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluVal 373
Db 1472 TCTTCTGTGACTCACAAACACCAATCCAAATCAAAATCAAAATCATGTGTACATTGATGAAGTT 1531
Qy 374 AlaGluAspProAlaLysSerLeuThrGluLysSerThrAspPheAspArgSerSerPro 393
Db 1532 GCAGAGACCCCTGCAAAATCTCTTACAGAGATATCTACAGACTTTGACCGTCTCTACCA 1591
Qy 394 ProLeuGlnProProValAsnSerLeuThrThrGluAsnArgPheHisSerLeuPro 413
Db 1592 CCACTCCAGCCTCCTCTGTGAACCTCACTGACCACCGAGAACAGATTCACACTCTTTACCA 1651
Qy 414 PheSerLeuThrLysMetProAsnThrAsnGlySerIleGlyHisSerProLeuSerLeu 433
Db 1652 TTCAGTCTCAACCAAGATGCCCAATACCAATGGAAGTATTGGCCACAGTCACACTTCTCTG 1711
Qy 434 SerAlaGlnSerValMetGluLeuLeuAsnThrAlaProValGlnGluSerProProLeu 453
Db 1712 TCAGCCAGTCTGTAAATGGAAGAGCTAAACACTGACCCCTGCAAGAGAGTCCACCCCTG 1771
Qy 454 AlaMetProProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValLysGlu 473
Db 1772 GCCATGGCTCTGGGAACCTCACATGGTCTAGAGTGGGCTCATTGGCTGAAGTTAAGGAG 1831
Qy 474 AsnProProPheTyrGlyValIleArgTrpIleGlyGlnProProGlyLeuAsnGluVal 493
Db 1832 AACCTCTCTTCTATGGGTAAATCCGTTGGATCGGTGATCCAGCCAGGACTGAATGAAGTG 1891
Qy 494 LeuAlaGlyLeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArgGly 513
Db 1892 CTCGCTGGACTGGAACCTGGAAGATGAGTGTGCAAGGCTGTACGGATGGAACCTTCAGAGGC 1951
Qy 514 ThrArgTyrPheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArg 533
Db 1952 ACTCGGTATTTCACTGTGCTGCTGGAAGAGCGCTGTTGTGAACCTGGAAGCTGCAGG 2011
Qy 534 ProAspSerArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsnSer 553
Db 2012 CCTGACTAGGTTTGACTCAATTCAGCCCGGTTTCCAAATCAGATTCAGCGCTGAACCTC 2071
Qy 554 LeuAlaPheGlyTyrLeuSerGluValValGluGluAsnThrProProLysMetGlu 573
Db 2072 TTAGCATTTGGAGGCTACTTAAGTGAAGTAGTAGAAGAAATATCTCCACCAAAATGGAA 2131
Qy 574 LysGluGlyLeuGluIleMetIleGlyLysLysLysGlyIleGlnGlyHisTyrAsnSer 593
Db 2132 AAAGAAGGCTGGAGATAATGATTTGGGAAGAAAGGCAATCCAGGCTCATTTCAATCT 2191
Qy 594 CysTyrLeuAspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThrVal 613
Db 2192 TGTTACTTAGACTCAACCTTATTCCTGTTATTTGCTTTTAGTTCTGTCTGGACACTGTG 2251
Qy 614 LeuLeuArgProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeu 633
Db 2252 TTACTTAGACCCCAAGAAAGACGATGATAGATATATTATAGTGAACCCCAAGAGCTACTG 2311
Qy 634 ArgThrGluIleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrLysIleMet 653
Db 2312 AGGACAGAAATTTTAATCCTCTGAGAATATATGATATGTGTGTGTCACAAAAATATG 2371
Qy 654 LysLeuArgLysIleLeuGluLysValGluAlaSerGlyPheThrSerGluLys 673
Db 2372 AAATCTGAGAAATACTCTGAAAGGTGGAGGCTGCATCAGGATTTACCTCTGAAGAAAA 2431
Qy 674 AspProGluGluPheLeuAsnIleLeuPheHisIleLeuArgValGluProLeuLeu 693
Db 2432 GATCCTGAGGAATCTTGAATATTTCTTTCATCATATTTTAAGGTGAGAACCTTTGCTA 2491

Qy 694 LysIleArgSerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGlu 713
Db 2492 AAAATAAGATCAGCAGGTCAAAAGGTCAAGATTGTTACTTCTATCAAAATTTTATGGAA 2551
Qy 714 LysAsnGluLysValGlyValProThrIleGlnGlnLeuLeuGluTrpSerPheIleAsn 733
Db 2552 AAAAATGAGAAGTTGGCTGCCACAATTCAGCAGTTGTTAGAAATGGCTTTTATCAAC 2611
Qy 734 SerAsnLeuLysPheAlaGluAlaProSerCysLeuIleIleGlnMetProArgPheGly 753
Db 2612 AGTAACCTGAATTTGAGAGGACCAATCATGTTCTGATTATTCAGATGCCCTCGATTGGA 2671
Qy 754 LysAspPheLysLeuPheLysLysIlePheProSerLeuGluLeuAsnIleThrAspLeu 773
Db 2672 AAAGACTTTAACTATTTAAAAAATTTTCTCTCTGGAATTTAAATATAACAGATTTA 2731
Qy 774 LeuGluAspThrProArgGlnCysArgIleCysGlyGlyLeuAlaMetTyrGluCysArg 793
Db 2732 CTTGAAGACACTCCACACAGTGCAGATATGTGGAGGCTTGCAAATGTATGAGTGTAGA 2791
Qy 794 GluCysTyrAspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCys 813
Db 2792 GAATGCTACACGATCCGACATCTCAGCTGGAATAATCAAGCAGTTTGTAAACACCTGC 2851
Qy 814 AsnThrGlnValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeu 833
Db 2852 AACACTCAAGTCCACCTTCATCCGAGAGGCTGAATCATTAATATAACCCAGTGTCACTT 2911
Qy 834 ProLysAspLeuProAspTrpAspTrpArgHisGlyCysIleProCysGlnAsnMetGlu 853
Db 2912 CCCAAAGACTTACCCGACTGGGACTGGAGACACGGCTGCATCCCTTGGCCAGAAATATGGAG 2971
Qy 854 LeuPheAlaValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLys 873
Db 2972 TTATTTGCTGTTCTCTGCTAGAAACAAAGCCACTATGTTGTTTTGTGAAGTATGGGAAG 3031
Qy 874 AspAspSerAlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlyGlnAsnGly 893
Db 3032 GACGATTCGCTGGCTCTCTTTTGACGATGCCGATCGGATGGGTGTCAGAAATGGC 3091
Qy 894 PheAsnIleProGlnValThrProCysProGluValGlyGluTyrLeuLysMetSerLeu 913
Db 3092 TTCAACATTCCTCAAGTCAACCCATGCCAGAGTAGGAGAGTACTTTGAAGATGTCTCTG 3151
Qy 914 GluAspLeuHisSerLeuAspSerArgArgIleGlnGlyCysAlaArgArgLeuLeuCys 933
Db 3152 GAAGACCTGCTATCTCTGGACTCCAGGAGAAATCCAGGCTGTGCACGAGACTGCTTTGT 3211
Qy 934 AspAlaTyrMetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949
Db 3212 GATGCATATATGTGCATGTACCAGAGTCCAACAATGATGTTGTACAA 3259
RESULT 4
BD231207
LOCUS 4527 bp DNA linear PAT 17-JUL-2003
DEFINITION Human cytoskeleton associated proteins.
ACCESSION BD231207
VERSION BD231207.1 GI:33040977
KEYWORDS JP 2002526076-A/9.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 4527)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
Lal, P., Tang, T. Y., Yue, H., Hillman, J. L., Bandman, O., Corley, N. C.,
Guegler, K. J., Patterson, C., Azimzai, Y. and Baughn, M. R.
TITLE Human cytoskeleton associated proteins
JOURNAL Patent: JP 2002526076-A 9 20-AUG-2002;
INCYTE PHARMACEUTICALS INC
OS Homo sapiens (human)
PN JP 2002526076-A/9
PD 20-AUG-2002

Db 1829 G C C C T G A A G A G G C G C T G T T T G T G A A A C T G A A G A G C T G C A G G C C T G A C T C T A G A G T T T G C A 1898
Qy 540 S e r L e u G l n P r o V a l S e r A s n G l n I l e u A r g C y s A s n S e r L e u A l a P h e G l y G l y T y r 559
Db 1889 T C A T T G C A G C G G T T C C A A T C A G A T T G A G G C T G T A A C T C T T T A G C A T T T G G A G G C T A C 1948
Qy 560 L e u S e r G l u V a l V a l G l u G l u A s n T h r P r o P r o L y s M e t G l u L y s G l u G l y L e u G l u I l e 579
Db 1949 T T A A G T G A A G T A G A A A A A T A C T C C A C A A A A A T G G A A A A A A A G A G G C T T G G A G A T A 2008
Qy 580 M e t I l e G l y L y s L y s G l y I l e G l n G l y H i s T y r A s n S e r C y s T y r L e u A s p S e r T h r 599
Db 2009 A T G A T T G G G A A G A A A A G C A T C A G G G C A T T A C A A T T C T T G T T A C T T A G A C T C A A C C 2068
Qy 600 L e u P h e C y s L e u P h e A l a P h e S e r S e r V a l L e u A s p T h r V a l L e u L e u A r g P r o L y s G l u 619
Db 2069 T T A T T C T G C T A T T T G C T T T A G T T C T G T T C T G A C A C T G T T A C T T A G A C C C A A A G A A 2128
Qy 620 L y s A s n A s p V a l G l u T y r T y r S e r G l u T h r G l n G l u L e u A r g T h r G l u I l e V a l A s n 639
Db 2129 A A G A A C G A T C T A G A A T A T T A T A G T G A A A C C A A G A G C T A C T G A G A C A G A A A T T G T T A A T 2188
Qy 640 P r o L e u a r g I l e T y r G l y T y r V a l C y s A l a T h r L y s I l e M e t L y s L e u A r g L y s I l e L e u 659
Db 2189 C C T C T G A G A T A T A T G A T A T T A T A G T G A A A C C A A G A G C T A C T G A G A C A G A A A T T G T T A A T 2248
Qy 660 G l u L y s V a l G l u A l a S e r G l y P h e T h r S e r C l u G l u L y s A s p P r o G l u G l u P h e L e u 679
Db 2249 G A A A A G T G A G G C T G C A T C A G A T T A C C T C T G A A A A A A A G A T C C T G A G G A A T T C T T G 2308
Qy 680 A s n I l e L e u P h e H i s I l e L e u A r g V a l G l u P r o L e u L y s I l e A r g S e r A l a G l y 699
Db 2309 A A T A T T C T G T T C A T C A T A T T T A A G G T A G A C C T T T G T A A A A A T A A G A T C A G C A G G T 2368
Qy 700 G l n L y s V a l G l n A s p C y s T y r P h e T y r G l n I l e P h e M e t G l u L y s A s n G l u L y s V a l G l y 719
Db 2369 C A A A A G T A C A A G A T T G T A C T C T A T C A A A T T T T A T G A A A A A A A T A G A A G T T G C 2428
Qy 720 V a l P r o T h r I l e G l n L e u L e u G l u T r P e r P h e I l e A s n S e r A s n L e u L y s P h e A l a 739
Db 2429 G T T C C C A C A T T C A G C A G T T G T T A G A T G T C T T T A T C A C A G T A A C C T G A A A T T T G C A 2488
Qy 740 G l u A l a P r o S e r C y s L e u I l e L e u G l n M e t P r o A r g P h e G l y L y s A s p P h e L y s L e u P h e 759
Db 2489 G A G C A C C A T C A T G T C G A T T A T T C A G A T C C T G A T T T G A A A A A G A C T T T A A A C T A T T T 2548
Qy 760 L y s L y s I l e P h e P r o S e r L e u G l u L e u A s n I l e T h r A s p L e u L e u G l u A s p T h r P r o A r g 779
Db 2549 A A A A A A T T T T C T C T C T G G A A T T A A A T A T A A C A G A T T A C T T G A A G A C A C T C C C A G A 2608
Qy 780 G l n C y s A r g I l e C y s G l y L e u A l a M e t T y r G l u C y s A r g G l u C y s T y r A s p A s p P r o 799
Db 2609 C A G T G C G G A T A T G T G A G G C T T G C A A T G A T G A G T G T A G A A A T G C T A C G A C A T C C G 2668
Qy 800 A s p I l e S e r A l a G l y L y s I l e L y s G l n P h e C y s L y s T h r C y s A n t h r G l n V a l H i s L e u 819
Db 2669 G A C A T C T C A C T G G A A A A T C A A G C A G T T T T G T A A A A C C T G C A A C A C T C A A G T C C A C T T 2728
Qy 820 H i s P r o L y s A r g L e u A s n H i s L y s T y r A s n P r o V a l S e r L e u P r o L y s A s p L e u P r o A s p 839
Db 2729 C A T C C G A A G A G G C T G A A T C A A T A A T A A C C A G T G C A C T T C C C A A A G A C T T A C C C G A C 2788
Qy 840 T r p A s p T r p A r g H i s G l y C y s I l e P r o C y s G l n A s n M e t G l u L e u P h e A l a V a l L e u C y s 859
Db 2789 T G G A C T T G G A G A C A C G C T G A T C C C T T G C C A G A T A T G A G T A T T T G C T G T C T C T G C 2848
Qy 860 I l e G l u T h r S e r H i s T y r V a l A l a P h e V a l L y s T y r G l y L y s A s p S e r A l a T r p L e u 879
Db 2849 A T A G A A A C A A G C C A C T A T G T T G C T T T G T G A A G T A T G G G A A G A C G A T T C G C T G C T C 2908
Qy 880 P h e P h e A s p S e r M e t A l a A s p A r g A s p G l y G l n A s n G l y P h e A s n I l e P r o G l n V a l 899

Db 2909 T T C T T T G A C A G C A T G C G C G A T G G T G G T C A G A A T G G C T T C A A C A T T C C T C A A G T C 2968
Qy 900 T h r P r o C y s P r o G l u V a l G l y L y s M e t S e r L e u L y s M e t S e r L e u G l u A s p L e u H i s S e r L e u 919
Db 2969 A C C C A T G C C C A G A A G T A G A G A G T A C T T T G A A G A T G T C T C T G A A A G A C C T G C A T T C C T T G 3028
Qy 920 A s p S e r A r g A r g I l e G l n G l y C y s A l a A r g A r g L e u L e u C y s A s p A l a T y r M e t C y s M e t 939
Db 3029 G A C T C C A G G A A T C C A A G G C T G T G C A C A A G A C T G C T T T G T A T G C A T A T A T G T G C A T G 3088
Qy 940 T y r G l n S e r P r o T h r M e t S e r L e u T y r L y s 949
Db 3089 T A C C A G A G T C C A A C A A T G A G T T T G T A C A A A 3118
RESULT 5
AKI22389 4314 bp mRNA linear ROD 15-MAR-2003
LOCUS Mus musculus mRNA for mKIAA0849 protein.
DEFINITION AKI22389
ACCESSION AKI22389.1 GI:28972434
VERSION FLI CDNA.
KEYWORDS Mus_musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Aizawa,H., Yuasa,S., Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
TITLE Prediction of the coding sequences of mouse homologues of KIAA gene: II. The complete nucleotide sequences of 400 mouse KIAA-homologous cDNAs identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries
JOURNAL DNA Res. 10, 35-48 (2003)
REFERENCE 2 (bases 1 to 4314)
AUTHORS Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
TITLE Direct Submission
JOURNAL Submitted (07-PEB-2003) Hisashi Koga, Kazusa DNA Research Institute, Laboratory for Genome Informatics; 2-6-7 Kazusa-kamakari, Kisarazu, Chiba 292-0818, Japan (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT The CREATE program supported by Japan science and technology corporation; cDNA full insert sequencing: Kazusa DNA Research Institute; cDNA library construction, clone selection and 5'- & 3'-end one pass sequencing.
FEATURES Location/Qualifiers
1..4314
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbh04492"
/tissue_type="brain"
/dev_stage="adult"
/note="vector:modified pBC SK+"
gene 1..4314
/gene="mKIAA0849"
<197..3196
/gene="mKIAA0849"
/note="CDS is predicted by in silico analysis. Start codon is not identified."
codon_start=1
evidence=not_experimental
/product="mKIAA0849 protein"
/protein_id="BAC65671.1"
/db_xref="GI:28972435"

NAHGLEVSGLAEVKNPPFYGVIRWIGQPPGLSDVLAGELEDEDECACTDGTFRGTRY
FTCALKALFVKLKSQRPSFASQPVSNQIERNCSLAFSGYGLSEVVEENTPPKMEK
EGLEIMKGGKIQGHYNSCYLDSTLFCFLFAPSSALDVLRLPKENKDI EYVSETOEL
LLETEVPLRLVGYVCAATKIMKRLKILEKVEASGFTSEKDPPEFLNILEHDLRLVE
PLTKRSAGQKQVDCNFQYIPEWKEKVGVPYTIQQLLEWSPFNKLEAPASCLIIQ
MPRFKDKLFKFIKPSLELNITLDLETPROCRICGGLAMYTECHRECYDDPDISAGKI
KQFCCTCTQVHLHRLNHSYHPVSLPKDLMDWRHGCTPCQWMEFLAVLICBTSH
YVAFVYKGDSDSAMLFFDSDMADRGGQNGFNIPQVTPCEVGEYKMSLEDLHLSLDR
RIQGCARRLLCCDAYMCWYQSPMTSLYK"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4314
Score: 4718.50 Matches: 898
Percent Similarity: 96.54% Conservative: 22
Best Local Similarity: 94.23% Mismatches: 28
Query Match: 93.73% Indels: 5
DB: 10 Gaps: 5

US-09-671-687A-3 (1-949) x AK122389 (1-4314)

QY 1 MetSerSerGlyLeuTrpSerGlnGluLysValThrSerProTyrTrpGluGluArgIle 20
DB 338 ATGAGTTTCAGGCCTGTGGAGCCCAAGAGAAAGTTACTTCAACCTACTGGGAAGACGGATT 397
QY 21 PheTyrLeuLeuGlnGluCysSerValThrAspLysGlnThrGlnLysLeuLeuLys 40
DB 398 TTTTATCTCTCTTCAAGAAATGCAAGTGAACAGACAAACAACTCAGAAAGCTGTGAAA 457
QY 41 ValProLysGlySerIleGlyGlnTyrIleGlnAspArgSerValGlyHisSerArgIle 60
DB 458 GTACCCAAAGGAGGATAGACAGTACATCCAAAGACCGTCTCTGTGGGGCAITCAAGATT 517
QY 61 ProSerAlaLysGlyLysLysGlnIleGlyLeuLysIleLeuGluGlnProHisAla 80
DB 518 CCTTCCACAAAAGGCAAGAAAATCAGATTGGATTAAAAATCTTGAGCAACCGCATGCA 577
QY 81 ValLeuPheValAspGlu---AspValValGluIleAsnGluLysPheThrGluLeuLeu 99
DB 578 GTTCTGTGTGTGATGAAAAGGATGTTGTAGAAAATAAATAAGAAAATTCACAGATTACTG 637
QY 100 LeuAlaIleThrAsnGlyGluArgPheSerLeuPheLysAsnArgAsnArgLeuSer 119
DB 638 TTGGCAATATACCACTGTGAGGAGAGGCTCAGCCTATTATTAGAAACAGACTCCGACTAAT 697
QY 120 LysGlyLeuGlnIleAspValGlyCysProValLysValGlnLeuArgSerGlyGluGlu 139
DB 698 AAAGGCCTCCAGGTAGACAGCTGGGCAGTCTGTGAAAGTACAGCTGCGATCTGGGAAAGAG 757
QY 140 LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGly 159
DB 758 AAATTTCCAGGAGTGTGTACGCTTCAGAGACCTTTATTAGCGGAGAGGACGGTGTGGGG 817
QY 160 IlePhePheGlyValGluLeuGluGlyValGlyArgGlyGlnGlyPheThrAspGlyVal 179
DB 818 ATTTCTTTTGGAGTAGAATATTGGNAGAAAGGTCGTGGTCAAGGTTTCACGATGGGGTA 877
QY 180 TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAsp 198
DB 878 TATCAAGGGAAGACAGCTTTTCCAGTGTGATGAGGACTGTGGCGTGTGTGTGATTTGGAC 937
QY 199 LysLeuGluLeuIleGluAspAspAspThrAlaLeuGluSerAspTyrAlaGlyProGly 218
DB 938 AAACATAGAACTTATAGAAGATGATGCAATGGAATTTGGAAGATGATTTTTCAGGCCCCAGGA 997
QY 219 AspThrMetGlnValGluLeuProLeuGluIleAsnSerArgValSerLeuLysGly 238
DB 998 GATACATACAGTTGAACTCTCCCTTTGGAAATAAATCTCCAGAGTTCTTTTGAAGGTT 1057
QY 239 GlyGluThrIleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSer 258
DB 1058 GGAGAAAGTACAGAACTCTGGAACAGTAATATCTGTGATGTTTACCAGGAAAAAGAGT 1117

QY 259 LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTrpAspGlyArgPhe 278
DB 1118 CTAGGATATTTTGTGGTGTGACATGATACCTCTATTGGCAACTGGGATGGAAGGTTT 1177
QY 279 AspGlyVal---LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsn 297
DB 1178 GATGGAGTACAGCTCTGTAGTTTGTCAAGTGTGTAAAGTACAAATCTCTCTGCACATCAAT 1237
QY 298 AspIleIleProGluSerValThrGlnGluArgArgProProLysLeuAlaPheMetSer 317
DB 1238 GACATCATCCAGATAGCTGACAGGAAAGAGGCTTCCCAAACTTGCTTTATGTCA 1297
QY 318 ArgGlyValGlyAspLysGlySerSerHisAsnLysProLysAlaThrGlySerThr 337
DB 1298 AGAGGTGTAGGTGACAAAGGTTTCATCTAGTATAATAAACCAAGGTTACAGGATCTACC 1357
QY 338 SerAspProGlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsnGlySerSerVal 356
DB 1358 TCAGACCCCTGGAAGTAGAAACAGATCTGAATTAATTTATACCTTAAATGGGTCTCTGTT 1417
QY 357 AspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluValAlaGluAsp 376
DB 1418 GACTCACAA---CAATCCAAAGTCCAAATAATCCATGATGATGAAGTTGCAGAGAC 1474
QY 377 ProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSerProProLeuGln 396
DB 1475 CTTGCRAAGTCACTTACAGAGATGTTCTCGGACTTCGGACATTCATCTCTCCACCGCAG 1534
QY 397 ProProProValAsnSerLeuThrGluAsnArgPheHisSerLeuProPheSerLeu 416
DB 1535 CCTCTCTCCATGAATCTCTGTGTAGCGAGAAACAGATTCATCTCTTACCTTTCAGCCTG 1594
QY 417 ThrLysMetProAsnThrAsnGlySerIleGlyHisSerProLeuSerLeuSerAlaGln 436
DB 1595 ACAAAGATGCCCAATACCTAATGGCAGCATGGGTCTAGTCCATCTCTCTGTGAGTGCAG 1654
QY 437 SerValMetGluGluLeuAsnThrAlaProValGlnGluSerProProLeuAlaMetPro 456
DB 1655 TCTGTGTGGGGAGGCTGNAACAGCACCTGTCCAGGAGAGTCCACCTTGCCCATCTCT 1714
QY 457 ProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValLysGluAsnProPro 476
DB 1715 TCTGGGAATGACACCGGCTAGAGGTGGGCTCCTGCTGGAAGTAAAGAGAAACCCCGG 1774
QY 477 PheTyrGlyValIleArgTrpIleGlyGlnProProGlyLeuAsnGluValLeuAlaGly 496
DB 1775 TTTCTATGGGGTTATCCGTTGGATTGGCCAGCCAGGAGCTCAGTGCAGTGTAGTGGGA 1834
QY 497 LeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArgGlyThrArgTyr 516
DB 1835 CTGGAACTGGAAGATGAATGCCAGGCTGTACAGATGGAACCTTTTACGGGGCACCGGTAT 1894
QY 517 PheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArgProAspSer 536
DB 1895 TTCAGCTGTGCTTGAAGAGGACCTGTTTGTGAAACTGAAGAGTGCAGACCGGACTCT 1954
QY 537 ArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsnSerLeuAlaPhe 556
DB 1955 AGGTTTGCATCTCTGCGAGCTGTTTCCAATCAGATTGAAAGGTGTAACCTTTTAGCATTT 2014
QY 557 GlyGlyTyrLeuSerGluValValGluGluAsnThrProProLysMetGluLysGluGly 576
DB 2015 GGGGGCTATTTAAGTGAAGTAGTAGAAGAAAATACTCCACCTTAAATGGAAGGAGGAGGT 2074
QY 577 LeuGluIleMetIleGlyLysLysGlyIleGlnGlyHisTyrAsnSerCysTyrLeu 596
DB 2075 TTAGAGATTAATGATTGGAAGAGAGGACATCCAGGCGCATTTACAACTTCTTGTACTTA 2134
QY 597 AspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThrValLeuLeuArg 616
DB 2135 GACTCAACTTTATCTGCTTATTTGCTTTAGTTCTGCGCTGACACACTGTGTTACTTAGA 2194
QY 617 ProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeuArgThrGlu 636

Db	2195	CCCAAAGAGAAGAAATGATATAGAGTATTACAGTGAGACTCAGGAGCTACTCAGGACACAG	2254
QY	637	IleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrLysIleMetLysLeuArg	656
Db	2255	ATAGTCAATCCTCTGAGAATATATGATATGTGTGTGCCAAGAATATTGAACACTGAGG	2314
QY	657	LysIleLeuGluLysValGluAlaAaSerGlyPheThrSerGluGluLysAspProGlu	676
Db	2315	AAAATACTTGAAAAGTTGAGGCTGCATCAGGTTTTACCTCTCGAAGAAAAGATCCTGAA	2374
QY	677	GluPheLeuAsnIleLeuPheHisIleLeuArgValGluProLeuLeuLysIleArg	696
Db	2375	GAATTCCTAAATATCCTGTTTCATGATATTTTAAGGGTTGAACCATTTGTTAAAAATAAGA	2434
QY	697	SerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAsnGlu	716
Db	2435	TCAGCAGGTCAAAAGTTCAAGACTGTAACTCTCATCAAAATTTTTATGGAAAAAATGAG	2494
QY	717	LysValGlyValProThrIleGlnGlnLeuLeuGluTrpSerPheIleAsnSerAsnLeu	736
Db	2495	AAAGTTGGAGTACCCACAATTCAGCAGTTATTAGAATGGTCTTTTATCAACAGCAACCTG	2554
QY	737	LysPheAlaGluAlaProSerCysLeuIleIleGlnMetProArgPheGlyLysAspPhe	756
Db	2555	AAATTTGCAGAGGCACCATCATGCTTGATTCAGATGCTCGGTTTGGAAAAAGACTTT	2614
QY	757	LysLeuPheLysIlePheProSerLeuGluLeuAsnIleThrAspLeuLeuGluAsp	776
Db	2615	AAACTATTTAAAAAAATTTTCTCCCTCGGAATTAATAATAACAGATTTACTTGAAGAC	2674
QY	777	ThrProArgGlnCysArgIleCysGlyGlyLeuAlaMetTyrGluCysArgGluCysTyr	796
Db	2675	ACTCCAGCGAGTCCCGCATCTGTGGAGACTTCCGATGACAGTGACGGAGTGCTAT	2734
QY	797	AspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAsnThrGln	816
Db	2735	GACGATCCGACATCTCAGCTGGAGAGATCAACAGCTTCTGTAAGACCTGCACACTCAG	2794
QY	817	ValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeuProLysAsp	836
Db	2795	GTTACCTTCATCCCAAGAGGTTGAATCATCTTATCATCATCCAGTATCACTTCCCAAAGAC	2854
QY	837	LeuProAspTrpAspTrpArgHisGlyCysIleProCysGlnAsnMetGluLeuPheAla	856
Db	2855	TTGCCTGACTGGGACTGGAGACACGGCTGCATCCCTGTGAGAAGATGGAGTTATTGCT	2914
QY	857	ValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspAspSer	876
Db	2915	GTTCTCTGATAGAAACGACCACTATGTGCTTTGTGAGTATGGGAAGATGACTCT	2974
QY	877	AlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlnAsnGlyPheAsnIle	896
Db	2975	GCGTGGCTTTCTTTGACAGCATGGCGGATCGAGATGGTGGTTCAGAAATGGCTTCAACATT	3034
QY	897	ProGlnValThrProCysProGluValGlyGluTyrLeuLysMetSerLeuGluAspLeu	916
Db	3035	CCACAAGTCACGCCCTGCCAGAAAGTGGGAGAGTACTTGAAGATGCTCTGGAGGACCTG	3094
QY	917	HisSerLeuAspSerArgGlnGlnGlyCysAlaArgArgLeuLeuCysAspAlaTyr	936
Db	3095	CACTCTTTGGATCTCAAGAGGATTCAGGCTGTGCGCGCAGACTCTTTTCCGATGCATAC	3154
QY	937	MetCysMetTyrGlnSerProThrMetSerLeuTyrLys	949
Db	3155	ATGTGCATGTACCAGAGTCCAAACCATGAGCCTGTACAAA	3193
RESULT 6			
BC042438			
LOCUS			
DEFINITION	BC042438	4501 bp	mRNA linear ROD 21-OCT-2003
ACCESSION	BC042438		Mus musculus cylindromatosis (turban tumor syndrome), mRNA (CDNA clone MGC:25429 IMAGE:3983771), complete cds.

VERSION
KEYWORDS
SOURCE

BC042438.1 GI:27503670

MGC.

Mus musculus (house mouse)

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 4501)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,K., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.C., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Shevchenko,Y.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Green,E.D.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

REFERENCE

2 (bases 1 to 4501)

Strausberg,R.

Direct Submission

Submitted (02-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgaps-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,

Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,

A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>

Series: IRAC Plate: 30 Row: g Column: 3.

Location/Qualifiers

1. .4501

/organism="Mus musculus"

/mol_type="mRNA"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="MGC:25429 IMAGE:3983771"

/tissue_type="Mammary tumor. Metallothionien-TGF alpha

model. 10 month old virgin mouse. Taken by biopsy."

/clone_lib="NCI CGAP_Mam1"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

1. .4501

/gene="CyclD"

/notes="synonyms: C130039D01Rik, 2900009M21Rik, EAC, CDMT,

CYLD1"

/db_xref="LocusID:74256"

/db_xref="MGI:1921506"

FEATURES
source

gene

CDS	176..3034 /codon_start=1 /product="RIKEN cDNA 2010013M14" /protein_id="AAH42438.1" /db_xref="GI:27503671" /db_xref="J:74256" /translation="MSSGLMSQEKVTPYWEERIFLYLLQBCSVTDKQTKLLKVPKG SIGQYIDRSVGHVRPSTKGKQNGIKLIEQPHAVLFVDEKQVVEINEKFTELL ITNCEERLSRPNRLSKGLQVDGSPVKQLRSGBEKFFPCVFRFGPLRPTVSG IFFGVELLEGRGGFTDGVYGGKQLFCODEDCGVFVALDKLELIEDDNGLESDPAG PGDTWQVPPPLEINSRVSLLKVGSTESGTVIFCDVLPKESLGFVFGVDMDNPIGNW DGRDGLQCSFASVESTILLIINDIIPDSVTOERRPPKLAFMRSRGVGDGSSSHNKP KVTGSDPGRNRSSELYTYLLNGSVDSQQSKNPNWYIDEVAEDPAKSLTEMSDDFG HSSPPPOPSMNSLSRPNFSLPTKMPNTNGSMAHSPSLSLSVQVMGELNSTPV QESPLPLTSSGNHGLEVSLEAVKENPFYGVIRWIGQPPGLSDVLAGLEDEECAG CTDGTFRGTRYFTCAKALKALFVKLSKCRPSRPFASLQPVSNQIBRCNSLAFGGYLSEV VEENTPKWEKEGLEIMIGKKKGIOGHYNSCLDSTLFLFAFSALDITVLLRPKERN DIETYSFEQLRETEIYNPLRIYGVYCATIKMKILKELEVAASGFTSEEDKPEEFL NILFHDILURVPLLRKIRGAKQKODCNFYQIFMEKNKGVPTTQQLLEWSPFINSLK FAEAPSLIIOMPRFKDFKPKIFPSLELNIITDLEDTPQCRICGGLAMYECEBC YDDPDISAGTKQKCKTCTOVHLHPRLNHSYHPVSLPKDLPMWMPHGICIPQCKME LFAVICIETSHYFAVKYKDDSAWLFPDSDMADRGGQGNIPQVTPCPPEVGEYLKM SLEDLSLDSRRIOGCARRLLCDAYMCMYQSPMTSLYK" 644..766 /note="CAP GLY; Region: CAP-Gly domain. CAP stands for cytoskeleton-associated proteins" /db_xref="CDD:pfam01302"	Db	596	AAATTTCCAGGAGTTGTACGCTTCAGAGACCTTTATTAGCGGAGAGACGGTGTCCGGG	655
		Qy	160	lIePhePheGlyValGluLeuLeuGluGluGlyArgGlyGlnGlyPheThrSerGlyVal	179
		Db	656	ATTTTCTTTGGAGTAGAATATTATGGAAGAAGTCGTGGTCAAGGTTTCACGATGGGTA	715
		Qy	180	TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAsp	198
		Db	716	TATCAAGGGAAGCAGCTTTTCCAGTGTGATGAGGACTGTGGCGTGTTCATTGCAATGGAC	775
		Qy	199	LysLeuGluLeuIleGluAspAspThrAlaLeuGluSerAspTyrAlaGlyProGly	218
		Db	776	AAACTAGAACTATTAGAAGATGATGACAAATGATGGAAGTGATTTTCAGGCCCCAGGA	835
		Qy	219	AspThrMetGlnValGluLeuProProLeuGluIleAsnSerArgValSerLeuLysGly	238
		Db	836	GATACAATGCAGGTTGAACCTCCCTCTTTGGAAATAAATCCAGAGATTTCTTTGAAGGTT	895
		Qy	239	GlyGluThrIleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSer	258
		Db	896	GGAGAAAGTACGAATCTGGAACAGTATATTCTGTGATGTTTTTACCAGAAAAGAGAGT	955
		Qy	259	LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTrpAspGlyArgPhe	278
		Db	956	CTAGGATATTTTGTGGTGTGGACATGATACCTCTATTGGCAACTGGGATGGAAGGTTT	1015
		Qy	279	AspGlyVal---LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsn	297
		Db	1016	GATGGAGTACAGCTCTGTAGTTTTTGCAGAGTGTGTAAGAGTACAAATCTCTCCACATCAAT	1075
		Qy	298	AspIleIleProGluSerValThrGlnGluArgArgProProLysLeuAlaPheMetSer	317
		Db	1076	GACATCATCCAGATGAGTGTGACAGGAAAGAGGCGCTCCCAAACTTGCTTTATGTCA	1135
		Qy	318	ArgGlyValGlyAspLysGlySerSerSerHisAsnLysProLysAlaThrGlySerThr	337
		Db	1136	AGAGGTGTAGGTGACAAAGTTTCATCTAGTCATAATAACCAAGAGTTACAGGATCTACC	1195
		Qy	338	SerAspProGlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsnGlySerSerVal	356
		Db	1196	TCAGACCCCTGGAAAGTAGAAACAGATCTGAATTAATTTTATACCTTAAATGGTTCATGTT	1255
		Qy	357	AspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluValAlaGluAsp	376
		Db	1256	GACTCACAA---CAATCCAAAGTCCAAAATCCATGGTACATTGATGAAGTTGCAGAGAC	1312
		Qy	377	ProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSerProProLeuGln	396
		Db	1313	CCTGCAAAAGTCACTTACAGAGATGTTCTCGGACTTCGGACATTCATCTCTCCACCGCAG	1372
		Qy	397	ProProProValAsnSerLeuThrThrGluAsnArgPheHisSerLeuProPheSerLeu	416
		Db	1373	CCTCTCTCCATGAACCTCTTGTGTAGCGAGAACAGATTCACATCTCTTACCTTCAGCCTG	1432
		Qy	417	ThrLysMetProAsnThrAsnGlySerIleGlyHisSerProLeuSerLeuSerAlaGln	436
		Db	1433	ACAAAGATGCCAAATACCTAATGCGACATGGGTCTATAGTCCACTCTCTCTGTGAGTCAG	1492
		Qy	437	SerValMetGluGluLeuAsnThrAlaProValGlnGluSerProProLeuAlaMetPro	456
		Db	1493	TCTGTGATGGGGAGACTGNACAGACACCTGTCCAGAGAGTCCACCTTGCCATCTCT	1552
		Qy	457	ProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValLysGluAsnProPro	476
		Db	1553	TCTGGGAATGCACACGGGTAGAGGTGGGCTCACTGGCTGAAGTAAAGAAACCCCGG	1612
		Qy	477	PheTyrGlyValIleArgTrpIleGlnProProGlyLeuAsnGluValLeuAlaGly	496
		Db	1613	TTCTATGGGGTTATCCGTTGGATTGGCCAGCCACGAGGCTCAGTACGCTGTAGTGGGA	1672
		Qy	497	LeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArgGlyThrArgTyr	516

ORIGIN

Alignment Scores:	
Pred. No.:	0
Score:	4718.50
Percent Similarity:	96.54%
Best Local Similarity:	94.23%
Query Match:	10
DB:	10
US-09-671-687A-3 (1-949) x BC042438 (1-4501)	
Qy	1 MetSerSerGlyLeuTrpSerGlnGluLysValThrSerProTyrTrpGluGluArgIle 20
Db	176 ATGAGTTCCAGGCTGTGGAGCCACAGAGAAAGTTACTTCACCTACTGGGAGAACGGATT 235
Qy	21 PheTyrLeuLeuGluGluCysSerValThrAspLysGlnThrGlnLysLeuLeuLys 40
Db	236 TTTTATCTGCTTCTTCAGAAATGCAAGTGTAAACAGACAAACAACTCAGAAAGCTGCTGAAA 295
Qy	41 ValProLysGlySerIleGlyGlnTyrIleGlnAspArgSerValGlyHisSerArgIle 60
Db	296 GTACCCAAAGGAGATAGGACAGTACATCCAAAGACCGTTCGTGGGGCAATCAAGAGTT 355
Qy	61 ProSerAlaLysGlyLysAsnGlnIleGlyLeuLysIleLeuGluGlnProHisAla 80
Db	356 CCTTCCAAAAGGCAAGAAAATCAGATTGATTAAAAATCTTGGAGCAACCGCATGCA 415
Qy	81 ValLeuPheValAspGlu---AspValValGluIleAsnGluLysPheThrGluLeuLeu 99
Db	416 GTTCTGTTTGTGATAAAGGATGTTGTAGAAATAAATGAAATAATTCACAGATTACTG 475
Qy	100 LeuAlaIleThrAsnCysGluGluArgPheSerLeuPheLysAsnArgAsnArgLeuSer 119
Db	476 TTGGCAATATACCACTGTGAGGAGAGGCTCACCTATTATTAGAAACAGACTCCGACTAAGT 535
Qy	120 LysGlyLeuGlnIleAspValGlyCysProValLysValGlnLeuArgSerGlyGluGlu 139
Db	536 AAAGGCTCCAGGTAGACGTGGGCAGTCTCTGTGAAAGTACAGCTGCGATCTGGGGAAGAG 595
Qy	140 LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGly 159

```
Db 1673 CTGGAATCGAAGATGAATCGCGCAGGCTGTACAGATGGAATCTTCAGGGGACCGCGTAT 1732
Qy 517 PheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArgProAspSer 536
Db 1733 TTCACGTGTGCCCTGAAGAGGACACTGCTTTGTGAAACTGAAGAGCTGCAGACCGGACTCT 1792
Qy 537 ArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsnSerLeuAlaPhe 556
Db 1793 AGGTTTGCATCTTGCAGCCTGTTTCCAAATGAAATGAAAGGTGTAATCTCTTTAGCATTT 1852
Qy 557 GlyGlyTyrLeuSerGluValValGluGluAsnThrProProLysMetGluLysGluGly 576
Db 1853 GGGGGCTATTAAAGTAGAGTAGAAGAAATACTCCACTAAATGGAAAGGAAGGT 1912
Qy 577 LeuGluIleMetIleGlyLysLysLysGlyIleGlnGlyHisTyrAsnSerCysTyrLeu 596
Db 1913 TTAGAGATAATGATTGGAAAGAGAAGAGGATCCAGGGCCATTACAAATCTTGTACTTA 1972
Qy 597 AspSerThrLeuPheCysLeuPheAlaPheSerValLeuAspThrValLeuLeuArg 616
Db 1973 GACTCAACTTTATTCTGCTTATTGCTTTTAGTTCTGCCCTGGACACTGTGTTACTTAGA 2032
Qy 617 ProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeuArgThrGlu 636
Db 2033 CCCAAAGAGAAGAATGATATAGAGTATTACAGTGAGACTCAGGAGCTACTGAGGACAGAG 2092
Qy 637 IleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrLysIleMetLysLeuArg 656
Db 2093 ATAGTCAATCTCTGAGATATATGATATGATGTGTGCCACAAGATTATGAAGTGGAG 2152
Qy 657 LysIleLeuGluLysValGluAlaSerGlyPheThrSerGluGluLysAspProGlu 676
Db 2153 AAAATCTTTGAAAGTTGAGGCTGCATCAGGTTTTTACCTCTGAAGAAAAAGATCTCTGAA 2212
Qy 677 GluPheLeuAsnIleLeuPheHisIleLeuArgValGluProLeuLeuLysIleArg 696
Db 2213 GAATCTTAAATATCTGTTTTCATGATATTTAAGGGTTGAACCATTTGTTAAAAATAAGA 2272
Qy 697 SerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAsnGlu 716
Db 2273 TCAGCAGGTCAAAAGTTCAAGACTGTAACTTCTATCAAAATTTTATGGAAAAAATAGAG 2332
Qy 717 LysValGlyValProThrIleGlnGlnLeuLeuGluTyrPheIleAsnSerAsnLeu 736
Db 2333 AAAAGTTGAGTACCCACAACTTACAGACTATTAGATGCTTTTATCAACAGCACTG 2392
Qy 737 LysPheAlaGluAlaProSerCysLeuIleIleGlnMetProArgPheGlyLysAspPhe 756
Db 2393 AAAATTTGCAGAGGCACCATGCTTGATTATCCAGATGCTCGGTTTGGAAAAAGACTTT 2452
Qy 757 LysLeuPheIlysIlePheProSerLeuGluLeuAsnIleThrAspLeuLeuGluAsp 776
Db 2453 AAACTATTAAAAAATTTTCTCTCCCTGGAATTAATAATAACAGATTTTACTTGAAGAC 2512
Qy 777 ThrProArgGlnCysArgIleCysGlyGlyLeuAlaMetTyrGluCysArgGluCysTyr 796
Db 2513 ACTCCAGCAGGAGTCCGCCATCTCTGGAGACTTCCGATGTACAGATGCAGGAGTACTAT 2572
Qy 797 AspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAsnThrGln 816
Db 2573 GACGATCCGGACATCTCAGCTGGGAAGATCAAGCAGTTCTGTGAAGACCTGCAGCACTCAG 2632
Qy 817 ValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeuProLysAsp 836
Db 2633 GTTCACTTTATCCAGGAGGTTGAATCAATCTTATCATCCAGTATCACTTCCCAAGAC 2692
Qy 837 LeuProAspThrAspTrpArgHisGlyCysIleProCysGlnAsnMetGluLeuPheAla 856
Db 2693 TTGCTTGACTGGGACTGGAGACACGCTGCTATCCCTGTGAGAGATGGAGTTATTGCT 2752
Qy 857 ValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspAspSer 876
Db 2753 GTTCTCTGATAGAAACGACCACTATGTTGCTTTTGTGAGTATGGGAAGATGACTCT 2812
```

```
Qy 877 AlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlyGlnAsnGlyPheAsnIle 896
Db 2813 GCCTGGCTTTTCTTTGACAGCATGGCGATCGAGATGCTGGTTCAGATGGCTTCAACATT 2872
Qy 897 ProGlnValThrProCysProGluValGlyGluTyrLeuLysMetSerLeuGluAspLeu 916
Db 2873 CCACAAGTGACGCCCTGCCAGAAAGTGGAGAGTACTTTGAAGATGCTCTCTGGAGGACCTG 2932
Qy 917 HisSerLeuAspSerArgAtgIleGlnGlyCysAlaArgArgLeuLeuCysAspAlaTyr 936
Db 2933 CACTCTTTGGATTCCAGAAGATTCAAGGCTGTGCGCGACACTTCTTTGCGATGCATAC 2992
Qy 937 MetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949
Db 2993 ATGTGCTATGATACCAGAGTCCAAACCATGAGCTGTACAAA 3031
```

```
RESULT 7
AX883937 2845 bp DNA linear PAT 17-DEC-2003
LOCUS Sequence 18842 from Patent EP1074617.
DEFINITION AX883937
ACCESSION AX883937
VERSION AX883937.1 GI:40038838
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota,T., Iwogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ihii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 18842 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
source Location/Qualifiers
1..2845
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
452..2647
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAE92170.1"
/db_xref="GI:40038839"
/translation="MQVELPPELEINSRVLSKVGETIISGTVIFCDVLPCKSLGYFVG
VMDNPIGNWDRFGDQVLCFAVCESTILLHINDIIPESVTQERRPPKLAFFMSRGVG
DKGSSHNKPKATGSDPGRNRSELFYTLNGSSVDSQSQSKQNTWYIDEVADPA
KSUTEISTDFDRSPPLQPPVNSLTTENRPHSLPFLTKMPTNGSIGHSPLSAQ
SVMEELNTAVQESPLAMPNGSHGLEVSLAEVKENPPFPYVIRWIGQPPGLNEVL
AGLEDECACTDGTFRGTRYFTCKALKKALFKVLCRSPDRFASLPQVSNQIERCN
SLAFGYLSEVSEENTPPMKESGLEIMICKGIQGHYNSCYLDSILFCLFAFSVL
DTVLPGKENDVEYSETOELLURTEIVNPLRIYGYCATKIMLRKILEKVEASGF
TSKEQPEEFNLTFPHILREVELLKRISAGQVQDCYFYQIMKEKQKVPTIQOL
LEWSFINSLKFAEAPSLIIQMPRFQKFKKIFPSLELNTIDLELLETPQCRIC
GGIAMECRCEYDDPDISAKIKQPCCTNTQVHLHPKRLNKNPVSLPKOLPDWDW
RHGCIJCONMELEFVLCIETSHYVAFVKYKDSAWLFEFDSMADRDGGONGFNIPQVT
PCPEVEYILKMSLEDLHLSRRRIQGCARLLCDAYMCMQSPHSLYK"
```

CDS

```
Alignment Scores:
Pred. No.: 0 Length: 2845
Score: 4601.00 Matches: 875
Percent Similarity: 99.32% Conservative: 0
Best Local Similarity: 99.32% Mismatches: 2
Query Match: 91.40% Indels: 4
DB: 6 Gaps: 4
US-09-671-687A-3 (1-949) x AX883937 (1-2845)
```

ORIGIN

```
Qy 73 LysIleLeuGluGlnProHisAlaValLeuPheValAspGlu---AspValValGluTle 91
Db 2 AAAATTCAGAGCAACCTCATGCGATGTTCTCTTTCTGATGAAAAGGATGTTGTAGAGATA 61
```

QY 92 AsnGluLysPheThrGluLeuLeuAlaIleThrAsnCysGluGluArgPheSerLeu 111
DB 62 AATGAAAAGTTTCACAGAGTTACTTTTGGCAATTACCAATTGTGAGAGAGGTTTCAGCGTG 121
QY 112 PheLysAsnArgAsnArgLeuSerLysGlyLeuGlnIleAspValGlyCysProValLys 131
DB 122 TTTTAAAAACAGAAAACAGACTAAGTAAGAGCCCTCCAAATAGACGTGGCGTGTCTGTGAAA 181
QY 132 ValGlnLeuArgSerClyGluGluLysPheProGlyValValArgPheArgGlyProLeu 151
DB 182 GTACAGCTGAGATCTGGGGAAGAAAATTTCTGGAGTTGTACGTTTCAGAGAGCCCGTG 241
QY 152 LeuAlaGluArgThrValSerGlyIlePhePheGlyValGluLeuLeuGluGluArg 171
DB 242 TTAGCAGAGAGACAGTCTCCGGAATATTTCTTTGGAGTTGAATTCCTGGGAAGAGTGTGT 301
QY 172 GlyGlnGlyPheThrAspGlyValTyrGlnGlyLysGlnLeuPheGlnCysAspGluAsp 191
DB 302 GGTCAAGGTTTCACTGACGGGTGTACCAAGGGAACAGCTTTTTCAGTGTGATGAAGAT 361
QY 192 CysGly----PheValAlaLeuAspLysLeuGluLeuIleGluAspAspThrAlaLeu 210
DB 362 TGTGGCGTGTGTTGTGCATTGGACAAGCTAGAACTCATAGAAGATGATGACACTGCAATG 421
QY 211 GluSerAspTyrAlaGlyProGlyAspThrMetGlnValGluLeuProProLeuGluIle 230
DB 422 GAAAGTGATTACGAGGTCCTGGGACACAAATGCAGGTCGAACTTCTCTCTTTGGAAATA 481
QY 231 AsnSerArgValSerLeuLysGlyGlyGluThrIleGluSerGlyThrValIlePheCys 250
DB 482 AACTCCAGAGTTCTTTTGAAGGTGGAGAAACAATAGAACTCTGGAACAGTTATATTCGT 541
QY 251 AspValLeuProGlyLysGluSerLeuGlyTyrPheValGlyValAspMetAspAsnPro 270
DB 542 GATGTTTTGCCAGGAAAAGAAAGCTTAGGATATTTTGTGGTGTGGACATGATGAATACCT 601
QY 271 IleGlyAsnTrpAspGlyVArgPheAspGlyVal---LeuCysSerPheAlaCysValGlu 289
DB 602 ATTGGCAACTGGGATGGAAGATTGATGGAGTGCAGCTTTGTAGTTTTCCGTGTGTGAA 661
QY 290 SerThrIleLeuHisIleAsnAspIleIleProGluSerValThrGlnGluArgArg 309
DB 662 AGTACAAATTCATTGCACATCAATGATATCATCCAGAGAGTGTGACGAGGAAGGAGG 721
QY 310 ProProLysLeuAlaPheMetSerArgGlyValGlyAspLysGlySerSerHisAsn 329
DB 722 CCTCCCAAACTTGCCTTTATGTCAAGAGGTGTGGGACAAAGGTTTCATCCAGTCATAAT 781
QY 330 LysProLysAlaThrClySerThrSerAspProGlyAsnArg---ArgSerGluLeuPhe 348
DB 782 AAACCAAAAGGCTACAGGATCTACCTCAGACCCCTGGAAATAGAAACAGATCTGAATATATT 841
QY 349 TyrThrLeuAsnGlySerSerValAspSerGlnProGlnSerLysSerLysAsnThrTrp 368
DB 842 TATACCTTAATGGGTCTTCTGTGTGACTCACAAACCACAAATCCAAATCAAAAAATACATG 901
QY 369 TyrIleAspGluValAlaGluAspProAlaLysSerLeuThrGluIleSerThrAspPhe 388
DB 902 TACATTGATGAAGTTGCAGAACCCCTGCAAAATCTCTTACAGAGATACTACAGACTTT 961
QY 389 AspArgSerSerProProLeuGlnProProProValAsnSerLeuThrThrGluAsnArg 408
DB 962 GACCGTTCTTACCACCACTCCAGCCTCTCTGTGTGAACCTCACTGACCACCCGAGAACAGA 1021
QY 409 PheHisSerLeuProPheSerLeuThrLysMetProAsnThrAsnGlySerIleGlyHis 428
DB 1022 TTCCACTCTTTACCAATTCAGTCTCACCAAGATGCCCAATACCNAATGGAAGTATTGGCCAC 1081
QY 429 SerProLeuSerLeuSerAlaGlnSerValMetGluGluLeuAsnThrAlaProValGln 448
DB 1082 AGTCCACTTTCTCTGTGAGCCAGTCTGTAATGGAAGAGCTAAACACTGCACCCCGTCCAA 1141
QY 449 GluSerProProLeuAlaMetProProGlyAsnSerHisGlyLeuValGlySerLeu 468

DB 1142 GAGAGTCCACCCCTTGGCCATGCTCTCTGGAACTCACATGGTCTAGAACTGGGCTCATTTG 1201
QY 469 AlaGluValLysGlyAsnProPheTyrGlyValIleArgTrpIleGlyGlnProPro 488
DB 1202 GCTGAAGTTAAGGAGAACCCCTCTTTTCTATGGGTAAATCCGTGGATCGGTGAGCCACCA 1261
QY 489 GlyLeuAsnGluValIleAlaGlyLeuGluLeuGluAspGluCysAlaGlyCysThrAsp 508
DB 1262 GGACTGAATGAAGTCTCGCTGGACTTGGAACTTGGAAAGATGAGTGTGCAGGCTGTACGGAT 1321
QY 509 GlyThrPheArgGlyThrArgTyrPheThrCysAlaLeuLysLysAlaLeuPheValLys 528
DB 1322 GGAACCTTCAGAGGACCTCGGTATTTTCACTGTGCGCTCAAGAGAGCGCTGTTGTGAAA 1381
QY 529 LeuLysSerCysArgProAspSerArgPheAlaSerLeuGlnProValSerAsnGlnIle 548
DB 1382 CTGAAGAGCTGCAGGCTGACTCTAGGTTTGCATCATTTGCAGCCGGTTTCCAATCAGATT 1441
QY 549 GluArgCysAsnSerLeuAlaPheGlyGlyTyrLeuSerGluValValGluGluAsnThr 568
DB 1442 GAGCGCTGTAACTCTTTAGCATTTTGAGGCTACTTAAAGTGAAGTAGTAGAAGAAATACT 1501
QY 569 ProProLysMetGluLysGluGlyLeuGluIleMetIleGlyLysLysLysGlyIleGln 588
DB 1502 CCACCAAAAATGGAAGAGAGGCTTGGAGATAATGATTGGGAGAGAAAGGCAATCCAG 1561
QY 589 GlyHisTyrAsnSerCysTyrLeuAspSerThrLeuPheCysLeuPheAlaPheSerSer 608
DB 1562 GGTCAATCAAAATCTTGTACTTAGACTCAACCTTATTTCTGCTTATTTGCTTTTAGTTCT 1621
QY 609 ValLeuAspThrValIleLeuLeuArgProLysGluLysAsnAspValGluTyrTyrSerGlu 628
DB 1622 GTTCTGGACACTGTGTACTTGGACCCAAAGAAAAGAACGATGAGAAATATTATAGTGAA 1681
QY 629 ThrGlnGluLeuLeuArgThrGluIleValAsnProLeuArgIleTyrGlyTyrValCys 648
DB 1682 ACCCAAGAGCTACTGAGNACAGAAATGTTTANTCTCTGAGAAATATATGATATGTGTGT 1741
QY 649 AlaThrLysIleMetLysLeuArgLysIleLeuGluLysValGluAlaAlaSerGlyPhe 668
DB 1742 GCCACAAAAATATGAAACTGAGGAAAAATACTTGAAGAGGTGGAGGCTGCATCAGGATTT 1801
QY 669 ThrSerGluGluLysAspProGluGluPheLeuAsnIleLeuPheHisIleLeuArg 688
DB 1802 ACCTCTGAAGAAAAAGATCTCTGAGGAATTTCTTGAATATTTCTGTTTCATCATATTTTAA 1861
QY 689 ValGluProLeuLeuLysIleArgSerAlaGlyGlnLysValGlnAspCysTyrPheTyr 708
DB 1862 GTAGAACCTTTCTCTAAAAATAAGATCAGCAGGTCAAAAGGTACAAATTTGTTACTTCTAT 1921
QY 709 GlnIlePheMetGluLysAsnGluLysValGlyValProThrIleGlnGlnLeuLeuGlu 728
DB 1922 CAAATTTTATCGAAAAAAATCGAAAAAGTTGGCGTTTCCCACAATTCAGCAGTTGTTAGAA 1981
QY 729 TrpSerPheIleAsnSerAsnLeuLysPheAlaGluAlaProSerCysLeuIleIleGln 748
DB 1982 TGGTCTTTTATCAACAGTAACCTGAAATTTGAGAGGACCATCATGTCTGTGATTATTCAG 2041
QY 749 MetProArgPheGlyLysAspPheLysLeuPheLysLysIlePheProSerLeuGluLeu 768
DB 2042 ATGCGCTCGAATTTGGAAAGACTTTAAACTATTTTAAAAAAATTTTTCTTCTCTGGAATTA 2101
QY 769 AsnIleThrAspLeuLeuGluAspThrProArgGlnCysArgIleCysGlyGlyLeuAla 788
DB 2102 AATATAACAGATTTTACTTGAAGACACTCCAGACAGTGCAGGATATGTGGAGGCTTGCA 2161
QY 789 MetTyrGluCysArgGluCysTyrAspAspProAspIleSerAlaGlyLysIleLysGln 808
DB 2162 ATGTATGAGTAGAGAATGCTACGACGATCCGGACATCTCAGCTGGAATAATCAAGCAG 2221
QY 809 PheCysLysThrCysAsnThrGlnValHisLeuHisProLysArgLeuAsnHisLysTyr 828

```

Db      2222 TTTTGTAAACCTGCAACACTCAAGTCCACCTTCATCCGAGAGGCTGAATCATATAATAT 2281
QY      829  AsnProValSerLeuProLysAspLeuProAspTyrArgHisGlyCysIlePro 848
Db      2282 AACCCAGTGTCACTTCCCAAGACTTACCCGACTGGGACTGGAGACAGCGCTGCATCCT 2341
QY      849  CysGlnAsnMetGluLeuPheAlaValLeuCysIleGluThrSerHisTyrValAlaPhe 868
Db      2342 TGCAGAAATAGAGTTAATTGCTGTTCTCTGCATAGAAACACAGCCACTATGTTGCTTTT 2401
QY      869  ValLysTyrGlyLysAspAspSerAlaTyrPhePheAspSerMetAlaAspAsp 888
Db      2402 GTGAAGTATGGGAAGACGATTCTGCTGCTCTCTTTCACAGCATGGCGATCGGGAT 2461
QY      889  GlyGlyGlnAsnGlyPheAsnIleProGlnValThrProCysProGluValGlyGluTyr 908
Db      2462 GGTGGTCAAGTGGCTTCAACATTCCTCAAGTACCCCATGCCAGAGAGTAGGAGATAC 2521
QY      909  LeuLysMetSerLeuGluAspLeuHisSerLeuAspSerArgGlnGlyCysAla 928
Db      2522 TTCAAGATGTCTCTGGAAGACCTGCATTCTCTTGACTCCAGGAGATCCCAAGGCTGTGCA 2581
QY      929  ArgArgLeuLeuCysAspAlaTyrMetCysMetTyrGlnSerProThrMetSerLeuTyr 948
Db      2582 CGAAGACTGCTTTGTGATGCATATATGTGCATGTACCAAGAGTCCAAACATGAGTTGTAC 2641
QY      949  Lys 949
Db      2642 AAA 2644

RESULT 8
BD160617 2845 bp DNA linear PAT 17-JAN-2003
LOCUS     Primer for synthesizing full-length cDNA and use thereof.
DEFINITION
ACCESSION BD160617.1 GI:27866375
VERSION    JP 2002191363-A/15460.
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2845)
AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE      Primer for synthesizing full-length cDNA and use thereof
JOURNAL    Patent: JP 2002191363-A 15460 09-JUL-2002;
            HELIX RESEARCH INSTITUTE
COMMENT    OS Homo sapiens (human)
            PN JP 2002191363-A/15460
            PD 09-JUL-2002
            PF 28-JUL-2000 JP 2000280990
            PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
            PI SAITO,
            PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
            PI KEIICHI NAGAI, TETSUJI OTSUKI
            PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/
            10,
            PC C12P21/02, C12Q1/68/C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
            Primer for synthesizing full-length cDNA and use thereof FH Key
            Location/Qualifiers
            (452)...(2644).
FEATURES             Location/Qualifiers
     source           1..2845
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
ORIGIN
Alignment Scores: 0 Length: 2845
Pred. No.: 4601.00 Matches: 875
Score: 99.32% Conservative: 0
Percent Similarity:

```

```

Best Local Similarity: 99.32% Mismatches: 2
Query Match: 91.40% Indels: 4
DB: 6 Gaps: 4
US-09-671-687a-3 (1-949) x BD160617 (1-2845)

QY      73  LysIleLeuGluGlnProHisAlaValLeuPheValAspGlu---AspValValGluIle 91
Db      2  AAAATTCTAGAGCAACCTCATGCAGTTCTCTCTTGTGATGAAAAGAGATGTTGTAGAGATA 61
QY      92  AsnGluLysPheThrGluLeuLeuAlaIleThrAsnCysGluGluArgPheSerLeu 111
Db      62  AATGAAAAGTTCACAGAGTTACTTTTGGCAATTACCAATTGTGAGGAGGTTTCAGCCTG 121
QY      112  PheLysAsnArgAsnArgLeuSerLysGlyLeuGlnIleAspValGlyCysProValLys 131
Db      122  TTTAAAAACAGAAACACAGACTAAGTAAAGGCTCCAAATAGACGTGGGTGTTCTGTGAAA 181
QY      132  ValGlnLeuArgSerGlyGluGluLysPheProGlyValValArgPheArgGlyProLeu 151
Db      182  GTACAGCTGAGATCTGGGGAAGAAAAATTTCTCGAGTTGTACGCTTCAGAGGACCCCTG 241
QY      152  LeuAlaGluArgThrValSerGlyIlePhePheGlyValGluLeuLeuGluGluGlyArg 171
Db      242  TTAGCAGAGAGGACAGTCTCCGGAATATTTCTTGGAGTTGAATTGCTGGAAGAGGTCT 301
QY      172  GlyGlnGlyPheThrAspGlyValTyrGlnGlyLysGlnLeuPheGlnCysAspGluAsp 191
Db      302  GGTCAAGGTTTCACTGACGGGGGTGTACCAAGGGAACAGACTTTTTCAGTGTGATGAAGAT 361
QY      192  CysGly---PheValAlaLeuAspLysLeuGluLeuIleGluAspAspThrAlaLeu 210
Db      362  TGTGGCGTGTGTTGTGATTTGGACAAGCTAGAACTCATAGAAGATGATGACACTGCAATG 421
QY      211  GluSerAspTyrAlaGlyProGlyAspThrMetGlnValGluLeuProProLeuGluIle 230
Db      422  GAAAGTGATTAACGACAGTCTCGGGACACAAATGAGGTCGAACTTCCTCTTGGAAATA 481
QY      231  AsnSerArgValSerLeuLysGlyGlyThrIleGluSerGlyThrValIlePheCys 250
Db      482  AACTCCAGAGTTCTTTTGAAGGTTGGAGAAACAATAGAACTTGGAAACAGTTATATTCTGT 541
QY      251  AspValLeuProGlyLysGluSerLeuGlyTyrPheValGlyValAspMetAspAsnPro 270
Db      542  GATGTTTTCAGGAAAGAAAGCTTAGGATATTTTGTGTGGACATGGATAACCT 601
QY      271  IleGlyAsnTrpAspGlyArgPheAspGlyVal---LeuCysSerPheAlaCysValGlu 289
Db      602  ATTTGGCAACTGGGATGGAAGATTGATGGAGTGCAGCTTTTGTAGTTTTCGCTGTTGAA 661
QY      290  SerThrIleLeuLeuHisIleAsnAspIleIleProGluSerValThrGlnGluArg 309
Db      662  AGTACAATTTCTATTGACATCAATGATATCATCCAGAGAGTGTGACACGAAAGAGG 721
QY      310  ProProLysLeuAlaPheMetSerArgGlyValGlyAspLysGlySerSerHisAsn 329
Db      722  CTTCCCAAACTGCTTTATGTCAAGAGGTGTGGGACAAAGGTTTCATCAGATCATAT 781
QY      330  LysProLysAlaThrGlySerThrSerAspProGlyAsnArg---ArgSerGluLeuPhe 348
Db      782  AAACCAAGGCTACAGGATCTACTCAGACCTCGGAANTAGAAACAGATCTGAATTATTT 841
QY      349  TyrThrLeuAsnGlySerSerValAspSerGlnProGlnSerLysSerLysAsnThrTrp 368
Db      842  TATACCTTAAATGGGTCTTCTGTTGACTCAACACCAATCCAAATCAAAAATACATGG 901
QY      369  TyrIleAspGluValAlaGluAspProAlaLysSerLeuThrGluIleSerThrAspPhe 388
Db      902  TACATTGATGAAGTTGCAGAGACCCCTGCAAAATCTCTTACAGAGATATCTACAGACTTT 961
QY      389  AspArgSerSerProProLeuGlnProProProValAsnSerLeuThrThrGluAsnArg 408
Db      962  GACCGTTCTTACCACCACTCCAGCCTCCTCTGTAACCTCACTGACCACCGAGACAGA 1021

```


QY 409 PheHisSerLeuProPheSerLeuThrLysMetProAsnThrAsnGlySerIleGlyHis 428
DB 1022 TTCACACTCTTACCAATTCAGTCTCACCAGATGCCCATACCAATGGAAGTATTGGCCAC 1081
QY 429 SerProLeuSerLeuSerAlaGlnSerValMetGluGluLeuAsnThrAlaProValGln 448
DB 1082 AGTCACACTTCTCTGTGACGCCAGTCTGTAATGGAAGAGCTAAACACATGCACCCCGTCCAA 1141
QY 449 GluSerProProLeuAlaMetProProGlyAsnSerHisGlyLeuGluValGlySerLeu 468
DB 1142 GAGAGTCCACCTCTGGCCATGCTCTCTGGAACTCACATGGTCTAGAAAGTGGGCTCATTG 1201
QY 469 AlaGluValLysGluAsnProProPheTyrGlyValIleArgTrpIleGlyGlnProPro 488
DB 1202 GCTGAAGTTAAGAGAACCTCTCTTCTAATGGGTAAATCCGTGGATCGGTGAGCCACCA 1261
QY 489 GlyLeuAsnGluValLeuAlaGlyLeuGluLeuGluAspGluCysAlaGlyCysThrAsp 508
DB 1262 GGACTGAAATGAAGTCTCGCTGACTGGAACTGGAAAGATGATGTGCAGGCTGTACGGAT 1321
QY 509 GlyThrPheArgGlyThrArgTyrPheThrCysAlaLeuLysLysAlaLeuPheValLys 528
DB 1322 GGAACCTTCAGAGGCACCTCGGTATTTCACTGTGCCCTGAAGAAGCGCTGTTGTGAA 1381
QY 529 LeuLysSerCysArgProAspSerArgPheAlaSerLeuGlnProValSerAsnGlnIle 548
DB 1382 CTGAAGAGCTGCAGGCCCTGACTGATAGTTTGATCATTCAGCCGGTTCCTCAATCAGATT 1441
QY 549 GluArgCysAsnSerLeuAlaPheGlyGlyTyrLeuSerGluValValGluGluAsnThr 568
DB 1442 GAGCCCTGTAACCTCTTACCATTTGAGGCTACTTAAGTGAAGTAGTAGAAGAAATACT 1501
QY 569 ProProLysMetGluLysGluGlyLeuGluIleMetIleGlyLysLysGlyIleGln 588
DB 1502 CCACCAAAAAATGAAAAAGAGGCTGGAGATAATGATTGGGAAGAAAGGATCCAG 1561
QY 589 GlyHisTyrAsnSerCysTyrLeuAspSerThrLeuPheCysLeuPheAlaPheSerSer 608
DB 1562 GGTCAATTACAATTTCTGTACTTAGACTCAACCTTATTCTCGCTTATTGCTTTTAGTTCT 1621
QY 609 ValLeuAspThrValLeuLeuArgProLysGluLysAsnAspValGluTyrTyrSerGlu 628
DB 1622 GTTCTGGACACCTGTGTTACTTGGACCCAAAGAAAGACGATGAGAAATATTATAGTGAA 1681
QY 629 ThrGlnGluLeuLeuArgThrGluIleValAsnProLeuArgIleTyrGlyTyrValCys 648
DB 1682 ACCCAAGAGCTACTGAGGACAGAAATTTGTAATCCTCTCAGATATATGATGTGT 1741
QY 649 AlaThrLysIleMetLysLeuArgLysIleLeuGluLysValGluAlaSerGlyPhe 668
DB 1742 GCCACAAAAATATTGAAACTGAGGAAAAATCTTGAAGGTGGAGCTGCATCAGGATTT 1801
QY 669 ThrSerGluLysAspProGluGluPheLeuAsnIleLeuPheHisIleLeuArg 688
DB 1802 ACCTCTGAAGAAAAAGATCCTCAGGAATTTCTTGAATATTCTGTTTCATCATATTTAAGG 1861
QY 689 ValGluProLeuLeuLysIleArgSerAlaGlyGlnLysValGlnAspCysTyrPheTyr 708
DB 1862 GTAGAACCTTTGCTAAAAAATAGATCAGCAGGTCAAAAGGTACAGATTGTTACTTCTAT 1921
QY 709 GlnIlePheMetGluLysAsnGluLysValGlyValProThrIleGlnLeuLeuGlu 728
DB 1922 CAAATTTTATGAAAAAATGAGAAAGTTGGGTTTCCCAATTCAGCAGTTGTTAGAA 1981
QY 729 TrpSerPheIleAsnSerAsnLeuLysPheAlaGluAlaProSerCysLeuIleGln 748
DB 1982 TGGTCTCTTTTATCAACAGTAGTAACCTGAAATTTGACAGAGGCCATCATGTCTGATTATTCAG 2041
QY 749 MetProAtqPheGlyLysAspPheLysLeuPheLysLysIlePheProSerLeuLeuLeu 768
DB 2042 ATGCCTCGAATTTGGAAAGACCTTTAACTATTATTAATAAATTTTTCCTCTCTGGAATTA 2101

QY 769 AsnIleThrAspLeuLeuGluAspThrProArgGlnCysArgIleCysGlyGlyLeuAla 788
DB 2102 AATATAACAGATTTACTTTGAGACACTCCAGACAGTGGCGGATATGTGGGGCTTGCA 2161
QY 789 MetTyrGluCysArgGluCysTyrAspAspProAspIleSerAlaGlyLysIleLysGln 808
DB 2162 ATGTATGAGTGTAGAGAAATGCTACGACGATCGGACATCTCAGCTGGAATAATCAAGCAG 2221
QY 809 PheCysLysThrCysAsnThrGlnValHisLeuHisProLysArgLeuAsnHisLysTyr 828
DB 2222 TTTTGTAAACCTTGCACACTCAAGTCCACCTTCATCCGAAGGGCTGAATCATATAATAT 2281
QY 829 AsnProValSerLeuProLysAspLeuProAspTyrAspTyrArgHisGlyCysIlePro 848
DB 2282 AACCCAGTGTCTACTTCCCAAGACCTTACCCGACTGGACTGGAGACACGGCTGCATCCCT 2341
QY 849 CysGlnAsnMetGluLeuPheAlaValLeuCysIleGluThrSerHisTyrValAlaPhe 868
DB 2342 TGCCAGAAATATGGAGTTATTTCGTCTCTCATAGAAACAAGCCACTATGTGCTTTT 2401
QY 869 ValLysTyrGlyLysAspAspSerAlaTrpLeuPhePheAspSerMetAlaAspArgAsp 888
DB 2402 GTGAAGTAGTGGAGAGACGATTCTGCTGGCTCTTCTTTGACAGCATGGCCGATCGGAT 2461
QY 889 GlyGlyGlnAsnGlyPheAsnIleProGlnValThrProCysProGluValGlyGluTyr 908
DB 2462 GGTGTGAGATGGCTTCAACATTCTCAAGTCAACCCATGCCCAAGTAGAGAGTAC 2521
QY 909 LeuLysMetSerLeuGluAspLeuHisSerLeuAspSerArgArgIleGlnGlyCysAla 928
DB 2522 TTGAAGATGTCTCTGGAGACCTGTCATTCCTTGGACTCCAGAGAAATCCAGGCTGTGCA 2581
QY 929 ArgArgLeuLeuCysAspAlaTyrMetCysMetTyrGlnSerProThrMetSerLeuTyr 948
DB 2582 CGAAGACTGCTTGTGTGTCATATATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 2641
QY 949 Lys 949
DB 2642 AAA 2644

RESULT 9
AK024348
LOCUS
DEFINITION
Homo sapiens cDNA FLJ14286 fis, clone PLACE1006129, highly similar to Homo sapiens HSPC057 mRNA.
ACCESSION
AK024348.1 GI:10436712
VERSION
oligo capping: fis (full insert sequence).
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1
Isogai T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Wagatsuma, M., Hosokiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T. NEDO human cDNA sequencing project
2 (bases 1 to 2845)
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction; 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center

etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES

source

Location/Qualifiers
1..2845
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLAC1006129"
/tissue_types="placenta"
/clone_lib="PLACE1"
/notes="cloning vector: pME18SFL3"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2845
Score: 4601.00 Matches: 875
Percent Similarity: 99.32% Conservative: 0
Best Local Similarity: 99.32% Mismatches: 2
Query Match: 91.40% Indels: 4
DB: 9 Gaps: 4

US-09-671-687A-3 (1-949) x AK024348 (1-2845)

QY 73 LysfLeuGluGlnProHisAlaValLeuPheValAspGlu---AspValValGluLe 91
DB 2 AAAATTCTAGAGAACCTCATGAGATTCTCTTTGTTGATGAAAGGATGTTGTAGAGATA 61
QY 92 AsnGluLysPheThrGluLeuLeuLeuAlaIleThrAsnCysGluGluAArgPheSerLeu 111
DB 62 ATGAAAAGTTCACAGAGTTACTTTTGGCAATTACCAATTGTGAGGAGAGTTTCAGCCTG 121
QY 112 PheLysAsnArgAsnArgLeuSerLysGlyLeuGlnIleAspValGlyCysProValLys 131
DB 122 TTTAAAACAGAAACAGACTAAGTAAAGGCTCCNAATAGACGTGGGCTCTCTGTGAAA 181
QY 132 ValGlnLeuArgSerGlyGluGluLysPheProGlyValValArgPheArgGlyProLeu 151
DB 182 GTACAGCTGAGATCTGGGGAAGAAAATTTCTGGAGTTGTACGCTTCAGAGGACCCCTG 241
QY 152 LeuAlaGluArgThrValSerGlyIlePhePheGlyValGluLeuGluGlyArg 171
DB 242 TTAGCAGAGAGGACGTCTCCGGAATATCTTTGGAGTTGAAATGCTTGGGAAGAGTCTGT 301
QY 172 GlyGlnGlyPheThrAspGlyValTyrgInGlyLysGlnLeuPheGlnCysAspGluAsp 191
DB 302 GGTCAAGTTTCTAGCGGGGTGTACCAAGGGAACAGACATTTTTCAGTGTGATGAAGAT 361
QY 192 CysGly---PheValAlaLeuAspLysLeuGluLeuIleGluAspAspThrAlaLeu 210
DB 362 TGTGGCGTGTGTTGTCATTGGACAAGCTAGAACTCATAGAAGATGATGACACTGCATTG 421
QY 211 GluSerAspTyraLaGlyProGlyAspThrMetGlnValGluLeuProProLeuGluLe 230
DB 422 GAAAGTGATTACGACAGTCTCGGGGACACAAATGAGTGCAGTGCAGCTTCTCTCTTTGGAATA 481
QY 231 AsnSerArgValSerLeuLysGlyGluThrIleGluSerGlyThrValIlePheCys 250
DB 482 AACTCCAGAGTTTCTTTGAAGGTTGGAGAAACAATAGAACTCTGGAACAGTTATATTCTGT 541
QY 251 AspValLeuProGlyLysGluSerLeuGlyTyraPheValGlyValAspMetAspAsnPro 270
DB 542 GATGTTTTCAGGAAAAGAAAGCTTAGGATATTTTGGTGTGACATGGATTAACCTT 601
QY 271 IleGlyAsnTrpAspGlyArgPheAspGlyVal---LeuCysSerPheAlaCysValGlu 289
DB 602 ATTGGCAACTGGGATGGAAGATTGATGAGTGCAGCTTGTAGTTTGGCTGTGTTGAA 661
QY 290 SerThrIleLeuLeuHisIleAsnAspIleIleProGluSerValThrGlnGluAArg 309
DB 662 AGTACAAATTTATGACATCATATGATATCATCCACAGAGTGTGACGAGGAAGAGG 721
QY 310 ProProLysLeuAlaPheMetSerArgGlyValGlyAspLysGlySerSerHisAsn 329

DB 722 CCTCCCAAACTTGCTTTATGTCAAGAGGTGTGGGGACAAAGGTTTCATCCAGTCATAAT 781
QY LysProLysAlaThrGlySerThrSerAspProGlyAsnArg---ArgSerGluLeuPhe 348
DB 782 AAACCAAGGCTACAGGATCTACCTCAGACCTCGAATATAGAAACAGATCTGAATATTTT 841
QY TyrThrLeuAsnGlySerSerValAspSerGlnProGlnSerLysSerLysSerLysThrTrp 368
DB 842 TATACCTTAAATGGGTCTTCTGTGTTGACTCAACAACCAATCAAAATCAAAATAATCATGG 901
QY TyrlleAspGluValAlaGluAspProAlaLysSerLeuThrGluIleSerThrAspPhe 388
DB 902 TACATTGATGAAGTTGCAGAAAGCCCTGCAAAATCTCTTACAGAGATATCTACAGACTTT 961
QY AspArgSerSerProProLeuGlnProProValAsnSerLeuThrThrGluAsnArg 408
DB 962 GACCGTTCTTACCAACCACTCAGCCCTCTCTGTGAACTCTACTGACCAACCGAAGAACAGA 1021
QY PheHisSerLeuProPheSerLeuThrLysMetProAsnThrAsnGlySerIleGlyHis 428
DB 1022 TTCCACTCTTTTACCATTTCAGTCTCACAAGATGCCAATACCAATGGAAGTATTGGCCAC 1081
QY 429 SerProLeuSerLeuSerAlaGlnSerValMetGluGluLeuAsnThrAlaProValGln 448
DB 1082 AGTCCACTTTCTCTGTGAGCCAGTCTGTAAATGGAAGAGCTAAACACTGCACCCGTCCAA 1141
QY 449 GluSerProProLeuAlaMetProProGlyAsnSerHisGlyLeuGluValGlySerLeu 468
DB 1142 GAGAGTCCACCTTGGCCATGCTCTCTGGGAACCTACATGTGTCTAGAAAGTGGGCTCATTTG 1201
QY 469 AlaGluValLysGluAsnProProPheTyrglyValIleArgTrpIleGlyGlnProPro 488
DB 1202 GCTGAAGTTAAGGAGAACCTCTCTTTCTATGGGGTATCCGTGGATCGGTGAGCCACA 1261
QY 489 GlyLeuAsnGluValLeuAlaGlyLeuGluLeuGluAspGluCysAlaGlyCysThrAsp 508
DB 1262 GGACTGAATGAAGTCTCGCTGGAGCTGGAACCTGGAAGATGAGTGTGAGGCTGTACGGAT 1321
QY 509 GlyThrPheArgGlyThrArgTyrgPheThrCysAlaLeuLysLysAlaLeuPheValLys 528
DB 1322 GGAACCTTTCAGAGCACTCGGTATTTTCACTGTCTCTGGAAGAGGCGCTGTTCGTGAAA 1381
QY 529 LeuLysSerCysArgProAspSerArgPheAlaSerLeuGlnProValSerAsnGlnIle 548
DB 1382 CTGAGAGAGCTGCAGCCCTGACTCTAGGTTCATTCATTCGAGCCGCTTCCATCAGATT 1441
QY 549 GluArgCysAsnSerLeuAlaPheGlyClyTyrgLeuSerGluValValGluGluAsnThr 568
DB 1442 GAGCGCTGTAACCTTTTAGCATTTGGAGGCTACTTAAGTGAAGTAGTAGAAGAAATACT 1501
QY 569 ProProLysMetGluLysGluGlyLeuGluIleMetIleGlyLysLysLysGlyIleGln 588
DB 1502 CCACCAAAATGGAAGAAAGGCTTGGAGATAATGATTGGGAAGAAAGGCAATCCAG 1561
QY 589 GlyHisTyraAsnSerCysTyrgLeuAspSerThrLeuPheCysLeuPheAlaPheSerSer 608
DB 1562 GGTCAATTACAAATTTCTGTACTTAGACTCAACCTTATTTCTCTTATTTGCTTTTATGTTCT 1621
QY 609 ValLeuAspThrValLeuLeuArgProLysGluLysAsnAspValGluTyraTyrgSerGlu 628
DB 1622 GTTCTGGACACTGTGTACTTGGACCCCAAGAAAGAACGATGTAGATAATATTATAGTGA 1681
QY 629 ThrGlnGluLeuLeuArgThrGluIleValAsnProLeuArgIleTyrgTyrgValCys 648
DB 1682 ACCCAAGAGCTACTGAGCAGCAAAATTTGTAATCTCTGAGAATATATGATATGTGTGT 1741
QY 649 AlaThrLysIleMetLysLeuArgLysIleLeuGluLysValGluAlaAlaSerGlyPhe 668
DB 1742 GCCACAAAAATTATGAACCTGAGGAAATACTTGAAAAGGTGGAGGCTGCATACGAGATT 1801
QY 669 ThrSerGluGluLysAspProGluGluPheLeuAsnIleLeuPheHisIleLeuArg 688
DB 1802 ACCTCTGAGAAAAGATCTCTGAGAAATTTCTTGAATATTTCTGTTTCATCATATTTAAGG 1861

```

Qy 689 ValGluProLeuLeuLysIleArgSerAlaGlyGlnLysValGlnAspCysTyrPheTyr 708
Db 1862 GTAGAACCTTTGCTAAAATAATAGATACAGAGGTCAAAAGGTACAGATGTTACTTCTAT 1921
Qy 709 GlnIlePheMetGluLysAsnGlnLysValGlyValProThrIleGlnLeuLeuGlu 728
Db 1922 CAAATTTTATGGAAAAAATGAGAAAGTGGCGTTCCACAAATTCAGCAGTTGTAGAA 1981
Qy 729 TrpSerPheIleAsnSerAsnLeuLysPheAlaGluAlaProSerCysLeuIleleGln 748
Db 1982 TGGTCTTTTATCAACAGTAACCTGAAATTTGCAGAGGCACCATCATGTCGTATTATTCAG 2041
Qy 749 MetProArgPheGlyLysAspPheLysLeuPheLysLysIlePheProSerLeuLeuLeu 768
Db 2042 ATGCTCTGATTTGGAAAGACTTTAAACTATTAAAAAATTTTCTCTCTGGAATTA 2101
Qy 769 AsnIleThrAspLeuLeuGluAspThrProArgGlnCysArgIleCysGlyLeuAla 788
Db 2102 AATAACAGATTTACTTTGAAGACACTCCAGACAGTCCGGATATGTGGAGGGCTTGCA 2161
Qy 789 MetTyrGluCysArgGluCysTyrAspAspProAspIleSerAlaGlyLysIleLysGln 808
Db 2162 ATGTATGAGTGTAGAGAATGCTACGACGATCCGACATCTCAGCTCGAAAAATCAAGCAG 2221
Qy 809 PheCysLysThrCysAsnThrGlnValHisLeuHisProLysArgLeuAsnHisLysTyr 828
Db 2222 TTTTGTAAACCTGCAACACTCAAGTCCACCTTCATCCGAAGAGCTGAATCATAAATAT 2281
Qy 829 AsnProValSerLeuProLysAspLeuProAspTTPAspTTPArgHisGlyCysIlePro 848
Db 2282 AACCCAGTGTACTTCCCAAGACTTACCGACTGGACTGGAGACACGGCTGCATCCCT 2341
Qy 849 CysGlnAsnMetGluLeuPheAlaValLeuCysIleGluThrSerHisTyrValAlaPhe 868
Db 2342 TGCCAGAATATGGAGTTATTGCTGCTCTCTCATAGAAACAAGCACCATTATGTTGCTTT 2401
Qy 869 ValLysTyrGlyLysAspSerAlaTrpLeuPhePheAspSerMetAlaAspArgAsp 888
Db 2402 GTGAAGTATGGGAAGGAGCATTTCTCGCTGGCTCTTTTTCAGCAGATGCCCATCGGGAT 2461
Qy 889 GlyGlyGlnAsnGlyPheAsnIleProGlnValThrProCysProGluValGlyGluTyr 908
Db 2462 GTGTGTGAGATGGCTTCAACATTTCTCAAGTCACCCATGCCAGATCCAGAGTAGAGATAC 2521
Qy 909 LeuLysMetSerLeuGluAspLeuHisSerLeuAspSerArgArgIleGlnGlyCysAla 928
Db 2522 TTGAAGATGCTCTGGAAGACCTGCATTCCTTGGACTCCAGGAGAAATCCAGGCTGTGCA 2581
Qy 929 ArgArgLeuLeuCysAspAlaTyrMetCysMetTyrGlnSerProThrMetSerLeuTyr 948
Db 2582 CGAAGACTGCTTTGTGATGCATATATGTGATGATGATGATGATGATGATGATGATGATG 2641
Qy 949 Lys 949
Db 2642 AAA 2644

RESULT 10
AC123449
LOCUS
DEFINITION Rattus norvegicus clone CH230-21D6, *** SEQUENCING IN PROGRESS ***,
3 unordered pieces.
ACCESSION AC123449
VERSION AC123449.3 GI:23265406
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 241990)
AUTHORS Muzny,D,Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Aisbrooks,S., Amin,A., Anguitano,D.,

```

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chaves, D., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuwa, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelimeh, O., Okwuonu, G., Olanpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Sherty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 241990)

Worley, K.C.

Direct Submission

Submitted (29-MAY-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 241990)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (21-SEP-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Sep 21, 2002 this sequence version replaced gi:21902787.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Aclab

(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the

sequence may extend beyond the ends of the clone and there may be

contigs that consist entirely of whole genome shotgun sequence

reads. Both end sequences and whole genome shotgun sequence only

contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

```

Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXOF
Center clone name: CH230-21D6
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 229402 bases at least Q40
Consensus quality: 231763 bases at least Q30
Consensus quality: 233057 bases at least Q20
Estimated insert size: 260790; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 238985: contig of 238985 bp in length
* 238986 239085: gap of unknown length
* 239086 240731: contig of 1646 bp in length
* 240732 240831: gap of unknown length
* 240832 241990: contig of 1159 bp in length.
* Location/Qualifiers
* 1..241990
  /organism="Rattus norvegicus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10116"
  /clone="CH230-21D6"
  1323..2205
    /note="clone boundary"
    clone_end:Sp6
    site:EcoRI
    end_sequence:BH269859"
  complement(236297..237124)
    /note="clone boundary"
    clone_end:T7
    site:EcoRI
    end_sequence:BH269858"

FEATURES
source
1..241990
misc_feature
1323..2205
  /note="clone boundary"
  clone_end:Sp6
  site:EcoRI
  end_sequence:BH269859"
misc_feature
complement(236297..237124)
  /note="clone boundary"
  clone_end:T7
  site:EcoRI
  end_sequence:BH269858"

ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 241990
Score: 4600.50 Matches: 883
Percent Similarity: 95.29% Conservative: 27
Best Local Similarity: 92.46% Mismatches: 38
Query Match: 91.39% Indels: 8
DB: 2 Gaps: 5

US-09-671-687A-3 (1-949) x AC123449 (1-241990)
QY 1 MetSerGlyLeuTrpSerGlnGluLeuValThrSerProTyrTrpGluGluArgIle 20
DB 205375 ATGAGTTTCAGGCTGTGGAAACCAAGAGAAAGTTACTTCAACCTACTGGGAAGAACCGCTT 205434
QY 21 PheTyrLeuLeuGlnGluCysSerValThrAspLysGlnThrGlnLysLeuLeuLys 40
DB 205435 TTTTATCTGCTTCTCAAGATGCAGTGTAAAC-----AAACAGACACAGAAAGCTCTCTAGA 205490
QY 41 ValProLysGlySerIleGlyInTyrIleGlnAspArgSerValGlyHisSerArgIle 60
DB 205491 GTACCCAAAGGAGCAGTAGACAGTACATCCAAGACTGTTTCCGTGGGCCATTCAGAGTT 205550
QY 61 ProSerAlaLysGlyLysAsnGlnIleGlyLeuLysIleLeuGlnProHisAla 80
DB 205551 CCTTCCTGCTAAAGGCAAGAAATATCAGATTGGATTAAATAATCTTAGAGCAACCGCATGCA 205610
QY 81 ValLeuPheValAspGlu---AspValValGluIleAsnGluLysPheThrGluLeuLeu 99

```

```

Db 205611 GTTCTGTTTGTGATGAAAGGATGTTCTAGAAAATAAATGAAAAATTCACAGAGTTACTGT 205670
QY 100 LeuAlaIleThrAsnCysGluGluArgPheSerLeuPheLeuAsnArgAsnArgLeuSer 119
DB 205671 TTGGCAATTACCAACTGTGAGGAGGCTGACGCTATTATTAGAAACAGAAATCCGACTAAGT 205730
QY 120 LysGlyLeuGlnIleAspValGlyCysProValLysValGlnLeuArgSerGlyGluGlu 139
DB 205731 AAAGGCTCCAGGTAGAGTGGGAGTCTCTGTGAGAGTACAGTCCGATCTGGGAGGAG 205790
QY 140 LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGly 159
DB 205791 AAGTTTCAGGAGTTGTACGCTTCAGAGGACCTTTATTAGCGAGAGGACGCTGTCCGGG 205850
QY 160 IlePhePheGlyValGluLeuLeuGluGluArgGlyGlnGlyGlnGlyPheThrAspGlyVal 179
DB 205851 ATTTCTTTGGAGTAGAATTACTGGAAAGAGTCTGTGCAAGGTTTCACTGATGGGGT 205910
QY 180 TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAsp 198
DB 205911 TATCAAGGAAACACAGCTCTTCAGTGTGATGAGGACTGTGGTGTGTTTGTTCATTTGAC 205970
QY 199 LysLeuGluLeuIleGluAspAspThrAlaLeuGluSerAspTyrAlaGlyProGly 218
DB 205971 AAGCTGGAGCTTATAGAAGATGATGACAATGGGTTGGAAAGTGATTTTGCAGGCCAGGA 206030
QY 219 AspThrMetGlnValGluLeuProLeuGluIleAsnSerArgValSerLeuLysGly 238
DB 206031 GATACAGTCCAGGTTGAACCTCCCTTTGAAAATAAACTCCAGAGTTTCTTTGAAGTT 206090
QY 239 GlyGluThrIleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSer 258
DB 206091 GGAGAAAGTACAGAAATCTGGACAGTGATATCTGTGATGTTTACCAGAGAAAGAGAGT 206150
QY 259 LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTrpAspGlyArgPhe 278
DB 206151 CTAGGATATTTTGTGTGGTGGACATGATGATAACCTATTGGCAACTGGGATGGAAGTTT 206210
QY 279 AspGlyVal---LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsn 297
DB 206211 GATGGGGTTTCAGCGTTTCAGTGTTCGCAAGTGTTCAGAGTACAGTTCTCTTACACATCAAT 206270
QY 298 AspIleIleProGluSerValThrGlnGluArgArgProProLysLeuAlaPheMetSer 317
DB 206271 GACATATCCAGATACGCGTGCACAGAAAGGAGACCTCCCAAACTTGCCTTTATGTCA 206330
QY 318 ArgGlyValGlyAspLysGlySerSerHisAsnLysProLysAlaThrGlySerThr 337
DB 206331 AGAGGTGTAGGTGACAAAGGCTCATCTAGTCATATAAACCAAGGTTACAGGATCTACC 206390
QY 338 SerAspProGlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsnGlySerSerVal 356
DB 206391 TCAGACCTCGAAGTAGAAGACAGATCTCAATTTATTTATATACCTTAAATGGGTCTCATCTGT 206450
QY 357 AspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluValAlaGluAsp 376
DB 206451 GACTCACACACAAATCCAAGTCCAAAACCCATGGTACATTGATGAAGTTGCAGAAAGAC 206510
QY 377 ProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSerProProLeuGln 396
DB 206511 CCTGCAAAAGTCATTACAGAGATGTCTTCAGACTTCGGACATTCATCGCTCCACCCGCA 206570
QY 397 ProProProValAsnSerLeuThrThrGluAsnArgPheHisSerLeuProPheSerLeu 416
DB 206571 CCTCTCTCCATGAATCCTTGTCTAGCGAGAACAGATTCCACTCTTACCCTCTCAGCCTG 206630
QY 417 ThrLysMetProAsnThrAsnGlySerIleGlyHisSerPro---LeuSerLeuSerAlaG 436
DB 206631 ACAAAAGATGCCAACCACTAATGGCAGCATGGCTCACAGTCCACTCTCTCTCTGTGAGTGC 206690
QY 436 lnSerValMetGluGluLeuAsnThrAlaProValGlnGluSerProProLeuAlaMetP 456

```

Db 206691 AGTCTGTGATGGGGAGCTGAACACAGCAGCCCTGTCAGGAGAGTCCACCCCTGTCGCCAGCT 206750
Qy 456 roProGlyAasnSerHisGlyLeuGluValGlySerLeuAlaGluValLysGluAasnProp 476
Db 206751 CTTCTGGGAATGGACACGGGCTAGAGGTGGGCTCACTGGCTGAAGTAAAGAGAACCCCC 206810
Qy 476 roPheTyrGlyValIleArgTfPilleGlyGlnProProGlyLeuAasnGluValLeuAlaG 496
Db 206811 CGTCTATGGGGTATCCGTGGATGGCCAGCCACAGGGCTCAGTCAGCTGCTCGCTG 206870
Qy 496 lyLeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArgGlyThrArgT 516
Db 206871 GATTGGAACCGGAAGATGAATCGCAGGTTCACGGATGGAATTTTCAGGGCACCGCT 206930
Qy 516 yrPheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArgProAspS 536
Db 206931 ATTTCCACCTGTGCCCTGAAGAAAGCACTGTTCTGTGAAACTGAAGAGCTGCACACAGACT 206990
Qy 536 erArgPheAlaSerLeuGlnProValSerAasnGlnIleGluArgCysAasnSerLeuAlaP 556
Db 206991 CTAGGTTTGACCTCTTGAGCCCTGTTTCCAATCAGATGAAAGGTGTAACCTCTTTAGCAT 207050
Qy 556 heGlyGlyTyrLeuSerGluValValGluGluAasnThrProProLysMetGluLysGluG 576
Db 207051 TTGGGGGCTACTTAAGTGAAGTAGTAGAAGAAATACGCCACCTAAATGGAAGAAAGGAG 207110
Qy 576 lyLeuGluIleMetIleGlyLysLysLysGlyIleGlnGlyHisTyrAasnSerCysTyrL 596
Db 207111 GTTTAGAGATATGATGGAAGAAAGAAAGGATCCAGGGCCATTAACAATTTCTGTACT 207170
Qy 596 euAspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThrValLeuA 616
Db 207171 TAGACTCAACTTATTTCTGCTTATTTGCTTTTAGTTCTGCCCTGACACTGTATTACTTA 207230
Qy 616 rgProLysGluLysAasnAspValGluTyrTyrSerGluThrGlnGluLeuLeuArgThrG 636
Db 207231 GACCCAAAGAGAAGATGACGTAGAGTATTACAGTGAGACTCAAGAGCTACTGAGGACAG 207290
Qy 636 luIleValAasnProLeuArgIleTyrGlyTyrValCysAlaThrIysIleMetLysLeuA 656
Db 207291 AGATAGTCAATCTCTGAGAATATATGATATGTGTGTGCCCAAGATATTGAAGCTGA 207350
Qy 656 rgLysIleLeuGluLysValGluAlaAAserGlyPheThrSerGluGluLysAspProG 676
Db 207351 GGAATAATCTTGAAAAGTTGAGGTGCATCAGGATTTACCTCTGAGGAAAAGATCCGTG 207410
Qy 676 luGluPheLeuAasnIleLeuPheHisIleAAserValGluProLeuLeuLysIleA 696
Db 207411 AAGAATTTCTAAACATCTGTTTCATGATATTTTAAAGGATGACCATTTGTTTAAAAATAT 207470
Qy 696 rgSerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAasnG 716
Db 207471 GGTGACGAGCTCAAAAAGTTCAAGACTGTAACCTTCTATCAAAATTTTATGGAAGAAATG 207530
Qy 716 luLysValGlyValProThrIleGlnGlnLeuLeuGluTrpSerPheIleAasnSerAasnL 736
Db 207531 AGAAAGTCGGAGTACCACAATCCAGCAGTTATTAGAATGTCTTTTATCAACAGCAACC 207590
Qy 736 euLysPheAlaGluAlaProSerCysLeuIleIleGlnMetProArgPheGlyLysAspP 756
Db 207591 TGAATTTTCGGAGGAGCCACCATATGCTTGATTTATCCAGATGCTTGGTTTGGAAAGACT 207650
Qy 756 heLysLeuPheLysIlePheProSerLeuGluLeuAasnIleThrAspLeuLeuGluA 776
Db 207651 TTAAACTATTTAAAAAATTTTTTCTCTCCCTGGAATTTAAATATGACAGATTTACTTGAAG 207710
Qy 776 spThrProArgGlnCysArgIleCysGlyLeuAlaMetTyrGluCysArgGluCyst 796
Db 207711 ACACCTCCAGGAGTCCCGCATCTGTGGAGGACTCCGCCATGTATGATGATGATGAGTGTAG 207770
Qy 796 yrAspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAasnThrG 816
Db 207771 ATGATGCCCGGACATCTCGGAGGGAAGATCAAGCAGTTCTGTAAAGACTCGCAGCATC 207830

Qy 816 lnValHisLeuHisProLysArgLeuAasnHisLysTyrAasnProValSerLeuProLysA 836
Db 207831 AGTTTCACTTCCAGAGACTGAATCACTTACCATCCAGTATCACTTCCCAAG 207890
Qy 836 spLeuProAspTrpAspTrpArgHisGlyCysIleProCysGlnAasnMetGluLeuPheA 856
Db 207891 ACTTCCCGACTGGGACTGGAGACATGCTCATCCCATGTCAGAAAGATGGAGTTATTG 207950
Qy 856 laValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspAspS 876
Db 207951 CTGTCTCTGCATGAACACAGCCACTATGTTGCTTTTGTGAAGTACGGGAAGGATGACT 208010
Qy 876 exAlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlyGlnAasnGlyPheAasnI 896
Db 208011 CTGCTTGGCTCTCTTTGACAGATGGCTGATCGAGATGGTGTGAGATGGCTTCAACA 208070
Qy 896 leProGlnValThrProCysProGluValGlyGluTyrLeuLysMetSerLeuGluAspL 916
Db 208071 TTCCACAGTAGACACCTCTCCAGAGTAGGAGAGTACTTGAAGATGTCTCTGAGGACC 208130
Qy 916 euHisSerLeuAspSerArgArgIleGlnGlyCysAlaArgArgLeuLeuCysAspAlaT 936
Db 208131 TGCACTCTTTGACTCCAGAGGATTCAGGCTGTGCGGCGACACTTCTTTGCGATGCAT 208190
Qy 936 yrMetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949
Db 208191 ACATGTGATGTACCAGAGTCCACCATGAGCTGTGACAA 208231
RESULT 11
AC098162/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-3014, *** SEQUENCING IN PROGRESS ***,
3 unordered pieces.
AC098162
AC098162.7 GI:30522215
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 251132)
Muzny, D.Marie., Metzker, M.Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Aisbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebrugergis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, X., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jacob, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Loulseghe, H., Lozard, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, N., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokeneme, O., Okwunonu, G., Olarunpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Popper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojias, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

COMMENT

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 251132
Score: 4600.50 Matches: 883
Percent Similarity: 95.29% Conservative: 27
Best Local Similarity: 92.46% Mismatches: 38
Query Match: 91.39% Indels: 8
DB: 2 Gaps: 5

US-09-671-687A-3 (1-949) x AC098162 (1-251132)

Qy 1 MetSerGlyLeuTrpSerGlnGluLysValThrSerProTyrTrpGluArgIle 20
Db 163209 ATGAGTTTCAGGCTGTGGACCAAGAGAAAGTTACTTCCCTACTGGGAGACGGCTT 163150
Qy 21 PheTyrLeuLeuGlnGluCysSerValThrAspLysGlnThrGlnLysLeuLys 40
Db 163149 TTTTATTCGCTTCTCAAGAAATGCAGTGTAAC----AAACAGACACAGAAAGTCTCTGAGA 163094
Qy 41 ValProLysGlySerIleGlyGlnTyrIleGlnAspArgSerValGlyHisSerArgIle 60
Db 163093 GTACCCAAAGGAGCATTAGGACAGTACATCAAGACTGTTCCGTGGGCGCATCAGAGTT 163034
Qy 61 ProSerAlaLysGlyLysLysAsnGlnIleGlyLeuLysIleLeuGluGlnProHisAla 80
Db 163033 CCTTCTGCTAAAGGCAAGAAAAATCAGATTGGATTAAAAATCTTAGAGCAACCCGATGCA 162974
Qy 81 ValLeuPheValAspGlu----AspValValGluLeuLeuLysPheThrGluLeuLeu 99
Db 162973 GTTCTGTTTGTGATGAAAAGGATGTTGTAGAAATAAATAAGAAAAATTCACAGAGTTACTG 162914
Qy 100 LeuAlaIleThrAsnCysGluGluArgPheSerLeuPheLysAsnArgAsnArgLeuSer 119
Db 162913 TTGGCAATTACCAACTGTGAGGAGAGGCTCAGCCCTATTAGAAACAGAAATCCGACTAGT 162854
Qy 120 LysGlyLeuGlnIleAspValGlyCysProValLysValGlnLeuArgSerGlyGluGlu 139
Db 162853 AAAGGCTCCAGGTAGACGTGGGCGAGTCTCTGTGAGAGTACAGTCCGATCTGGGAGGAG 162794
Qy 140 LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGly 159
Db 162793 AAGTTTCAGGAGTTGTACGCTTCAGAGGACCTTTATTAGCGGAGAGACGCTGTCTGGGG 162734
Qy 160 IlePhePheGlyValGluLeuLeuGluGluClyArgGlyGlnGlyPheThrAspGlyVal 179
Db 162733 ATTTTCTTTGGAGTAGAAATTTACTGGAAGAAGGTCTGTGGCCCAAGGTTTTCATGATGGGGT 162674
Qy 180 TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAsp 198
Db 162673 TATCAGAGAAAACAGCTCTTCCAGTGTGATGAGAGACTGTGGTGTGTTTGTTCATTGGAC 162614

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GGWL

Center clone name: CH230-3014

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 241087 bases at least Q40

Consensus quality: 243305 bases at least Q30

Consensus quality: 244914 bases at least Q20

Estimated insert size: 256850; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

- * NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
- * NOTE: This sequence may represent more than one clone.
- * NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

Qy 199 LysLeuGluLeuIleGluAspAspThrAlaLeuGluSerAspTyrAlaGlyProGly 218
Db 162613 AAGCTGGAGCTTATAGAAAGATGATCAACATGGTTGGAAGATGATTTTGCAGGCCCGAGA 162554
Qy 219 AspThrMetGlnValGluLeuProProLeuGluIleAsnSerArgValSerLeuLysGly 238
Db 162553 GATACAGTCCAGGTTGAACCTCCCTTTGGAAATAACTCCAGAGTTTCTTTGAAGGTT 162494
Qy 239 GlyGluThrIleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSer 258
Db 162493 GGAGAAAGATACAGAACTCGGAACAGTATCTCTGTGATGTTTACCGAGGAAAGAGAGT 162434
Qy 259 LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTrpAspGlyArgPhe 278
Db 162433 CTAGGATATTGTGTGGTGGACATGATACCTATTGGCACTGGGATGGGAAGGTTT 162374
Qy 279 AspGlyVal---LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsn 297
Db 162373 GATGGGTTTCAGCGTTGCAGTTTTCGAAGTGTGTGAGAGTACAGTTCTCTACACATCAAT 162314
Qy 298 AsptIleIleProGluSerValThrGlnGluArgArgProProLysLeuAlaPheMetSer 317
Db 162313 GACATCATCCAGATAGCGTGACACAGGAAAGAGACCTCCCAAACTTGCCTTTATGTCA 162254
Qy 318 ArgGlyValGlyAspLysGlySerSerHisAsnLysProLysLeuAlaThrGlySerThr 337
Db 162253 AGAGGTGTAGGTGACAAAGGCTCATCTAGTCATAATAACCAAGGTTACAGGATCTACC 162194
Qy 338 SerAspProGlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsnGlySerVal 356
Db 162193 TCAGACCTCGNAGTAGAAACAGATCTGNAATTATTTTATACCTTAAATGGGTCTCTGT 162134
Qy 357 AspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluValAlaGluAsp 376
Db 162133 GACTCACAAACAATCAAGTCCAAAAACCCTCGTACATTGATGAAGTTCGAGAAAGAC 162074
Qy 377 ProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSerProLeuGln 396
Db 162073 CCTGCAAAAGTCACTTACAGAGATGCTTCAGACTTCGAGCATTCATCGCCTCCACCGCAA 162014
Qy 397 ProProValAsnSerLeuThrThrGluAsnArgPheHisSerLeuProPheSerLeu 416
Db 162013 CCTCTTCCATGAATCTCTGTCTAGCGAAGAACAGATTCCATCTCTTACCTTTCAGCCTG 161954
Qy 417 ThrLysMetProAsnThrAsnGlySerIleGlyHisSerPro--LeuSerLeuSerAlaG 436
Db 161953 ACAAGATGCCCAACACTAATGGCAGCATGGCTCACAGTCCACTCTCTCTGTCTGTCAGTGC 161894
Qy 436 InSerValMetGluGluLeuAsnThrAlaProValGlnGluSerProProLeuAlaMetP 456
Db 161893 AGTCTGTATGGGGAGCTGAAACAGCAGCGCTGTCCAGAGAGTCCACCTTTGCCCGAGCT 161834
Qy 456 roProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValLysGluAsnProp 476
Db 161833 CTTCTGGGAATGGACACGGGCTAGAGTGGGCTCACTGGCTGAGTAAAGAGAACCCCC 161774
Qy 476 roPheTyrGlyValIleArgTrpIleGlyGlnProProGlyLeuAsnGluValLeuAlaG 496
Db 161773 CGTCTATGGGTTTATCCGTTGGATTGGCCACGCCACCGGCTCAGTGACGTGCTGCTG 161714
Qy 496 lYLeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArgGlyThrArgT 516
Db 161713 GATTGAAACCGGAAGATGAATCGCAGGTTGACCGGATGGAACTTTTACGGGGCAGCGCT 161654
Qy 516 yrPheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArgProAspS 536
Db 161653 ATTTCCCTGTGCTTGAAGAAAGCACTGTTTCTGTAACCTGAAGACTGCAGACCCAGCT 161594
Qy 536 erArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsnSerLeuAlaP 556
Db 161593 CTAGGTTTGCATCTTTCAGCGCTGTTTCCAAATCAGATCGAAAGGTGTAATCTTTAGCAT 161534

Qy 556 heGlyGlyTyrLeuSerGluValValGluGluAsnThrProProLysMetGluLysGluG 576
Db 161533 TTGGGGCTACTTAACTGAGTAGTAGAAGAAATACGCCACCTAAATATGGAAGAGAG 161474
Qy 576 lYLeuGluIleMetIleGlyLysLysGlyIleGlnGlyHisTyrAsnSerCysTyrL 596
Db 161473 GTTTAGAGATAATGATTGGAAAGAAAGGATCCAGGGCCATTACAATTTCTTGTACT 161414
Qy 596 euAspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThrValLeuA 616
Db 161413 TAGACTCAACTTATTTCTGCTTATTTGCTTTTAGTTTCTGCCCTGGACACTGATTACTTA 161354
Qy 616 rgProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeuArgThrG 636
Db 161353 GACCCAAAGAGAGAAATGACGTAGAGTATTACAGTGAGACTCAAGAGCTACTGAGGACAG 161294
Qy 636 luIleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrIleMetLysLeuA 656
Db 161293 AGATAGTCAATCTCTGAGATATATGGATATGTGTGTGCCCAAAAGATTATGAAGCTGA 161234
Qy 656 rgLysIleLeuGluLysValGluAlaAlaSerGlyPheThrSerGluGluLysAspProG 676
Db 161233 GGAAATATCTTGAAGAAAGTTGAGGCTGCATCAGGATTTTACCTCTGAGGAAAGATCCTG 161174
Qy 676 luGluPheLeuAsnIleLeuPheHisIleLeuArgValGluProLeuLeuLysIleA 696
Db 161173 AAGAAATTTCTAAACATCTCTGTTTCATGATATTTTAAAGGATGACCATTTGTTTAAAAATAT 161114
Qy 696 rgSerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAsnG 716
Db 161113 GGTTCAGCAGGTCAAAAAGTTCAAGACTGTAACTTCTATCAAAATTTTATGGAAAAAATG 161054
Qy 716 luLysValGlyValProThrIleGlnGlnLeuLeuGluTrpSerPheIleAsnSerAsnL 736
Db 161053 AGAAAGTCGGAGTACCCCAATCCAGCAGATTATTAAGATTTCTTTTATCAACAGCAACC 160994
Qy 736 euLysPheAlaGluAlaProSerCysLeuIleIleGlnMetProArgPheGlyLysAspP 756
Db 160993 TGAAATTTTCGGAGGAGCCATCATCTGTTGATTATCCAGATGCTTGGTTTGGAAAGACT 160934
Qy 756 heLysLeuPheLysLysIlePheProSerLeuGluLeuAsnIleThrAspLeuGluA 776
Db 160933 TTAAACTATTTAAAAAATTTTTTCTCTCGAATTAATAATATGACAGATTATCTTGAAG 160874
Qy 776 spThrProArgGlnCysArgIleCysGlyLeuAlaMetTyrGluCysArgGluCysT 796
Db 160873 ACACCTCCAGGAGTCCCGCATCTGTGGAGGACTCCCATGATGATGATGATAGAGTGCT 160814
Qy 796 yrAspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAsnThrG 816
Db 160813 ATGATGACCCGAGCATCTCGGACGGGAAGATCAAGCAGTTCTGTAGACCTGCAGCACTC 160754
Qy 816 InValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeuProLysA 836
Db 160753 AGTTTCACTTTCATCCAGAAAGACTGAATCAACACTTACCATTCCAGTATCACTTCCCAAAG 160694
Qy 836 spLeuProAspTrpAspTrpArgHisGlyCysIleProCysGlnAsnMetGluLeuPheA 856
Db 160693 ACTTGGCCGACTGGGACATGGCTGCATCCCATGTCAGAAAGATGGAGTTATTG 160634
Qy 856 laValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspAspS 876
Db 160633 CTGTGCTCTGCATAGAAACACGCCACTATGTTGCTTTTGTGAAGTACCGGAAGGATGACT 160574
Qy 876 erAlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlyGlnAsnGlyPheAsnI 896
Db 160573 CTGCTCTGGCTCTTCTTTGACAGCATGGCTGATCGAGATGGTGTGCGAATGGCTTCAACA 160514
Qy 896 leProGlnValThrProCysProGluValGlyLysTyrLeuLysMetSerLeuGluAspL 916
Db 160513 TTCCCAAGTGCACCCCTCCAGAGAGTAGGAGTACTTGAAGATGTCTCTCTGGAGACC 160454
Qy 916 euHisSerLeuAspSerArgIleGlnGlyCysAlaArgArgLeuLeuCysAspAlaT 936

```
Db 160453 TGCACTCTTTGGACTCCAGAGGATTCAAGGCTGTGGCGCAGACTTCTTTGGCATGCAT 160394
QY 936 YrMetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949
Db 160393 ACATGTGCATGTACAGAGTCCACCACTGAGCTTGTACAAA 160353

RESULT 12
AR338799
LOCUS AR338799 2523 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 290 from patent US 6569662.
ACCESSION AR338799
VERSION AR338799.1 GI:33725656
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2523)
Tang,Y.F., Zhou,P. and Drmanac,R.T.
TITLE Nucleic acids and polypeptides
JOURNAL Patent: US 6569662-A 290 27-MAY-2003;
FEATURES
Location/Qualifiers
source 1..2523
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 2,9e-277 Length: 2523
Score: 3676.00 Matches: 692
Percent Similarity: 99.28% Conservative: 0
Best Local Similarity: 99.28% Mismatches: 3
Query Match: 73.02% Indels: 2
DB: 6 Gaps: 2

US-09-671-687A-3 (1-949) x AR338799 (1-2523)
QY 255 GlyLysGluSerLeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTyr 274
Db 45 GGAATAATAAGCTTTTATATTTTGTGGTGGACATGATTAACCTTATGGCAACTGG 104
QY 275 AspGlyArgPheAspGlyVal---LeuCysSerPheAlaCysValGluSerThrIleLeu 293
Db 105 GATGGAAGATTTCATGAGGTGCAGCTTTGTAGTTTTCGCGTGTGTGAAGTACAAATCTA 164
QY 294 LeuHisIleAsnAspIleIleProGluSerValThrGlnGluArgArgProProLysLeu 313
Db 165 TTGCACATCAATGATATCATCCAGAGAGTGTGACGCGAGAAAGAGGAGGCTCCCAACTT 224
QY 314 AlaPheMetSerArgGlyValGlyValGlyAspLysGlySerSerHisIleAsnLysProLysAla 333
Db 225 GCCTTTATGTCAAGAGGTGTGGGGCAAAAGGTTTCATCCAGTCATATAAACAAGGCT 284
QY 334 ThrGlySerThrSerAspProGlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsn 352
Db 285 ACAGGATCTACCTCAGACCTCGAATATAGAAACAGATCTGAATATTTTATACCTTAAT 344
QY 353 GlySerSerValAspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGlu 372
Db 345 GGGTCTCTGTTGACTCACAAACACCAATCCAAATCAAAATAATACATGTTACATTGATGAA 404
QY 373 ValAlaGluAspProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSer 392
Db 405 GTTGCAGAGACCCCTGCAAAATCTCTTACAGAGATATCTACAGACTTTTGACCGTCTTCA 464
QY 393 ProProLeuGlnProProValAsnSerLeuThrThrGluAsnArgPheHisSerLeu 412
Db 465 CCACCACTCCAGCTCTCTCTGTGAATCTACTGACCAACCCAGAACAGATTCCACTCTTTA 524
QY 413 ProPheSerLeuThrLysMetProAsnThrAsnGlySerIleGlyHisSerProLeuSer 432
Db 525 CCATTTCAGTCTCAACCAAGATGCCCAATACCAATGGGAAGTATTGGCCACAGTCCACTTCT 584
```

```
QY 433 LeuSerAlaGlnSerValMetGluGluLeuAsnThrAlaProValGlnGluSerProPro 452
Db 585 CTGTCAAGCCAGTCTGTAAATGGAAGAGCTAAACACTGCACCGTCCAAGAGAGTCCACCC 644
QY 453 LeuAlaMetProProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValLys 472
Db 645 TTGGCCATGCGCTCTGGGAACCTCATGGTGTAGAAAGTGGGCTCATTTGGCTGAAGTTAAG 704
QY 473 GluAsnProProPheTyrGlyValIleArgTrpIleGlyGlnProProGlyLeuAsnGlu 492
Db 705 GAGAACCTCTCTTCTATGGGTAATCCGTGGATCGGTGAGCCACAGGACTGAAATGAA 764
QY 493 ValLeuAlaGlyLeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArg 512
Db 765 GTGCTCGCTGAGCTGGAACCTGGAAGATGAGTGTGAGGCTGTACGGATGGAACCTTCAGA 824
QY 513 GlyThrArgTyrPheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCys 532
Db 825 GGCACCTCGGTATTTTCACTGTGCCCTGAAGAGGCGCTGTTTGTGAAACTGAAGAGCTGC 884
QY 533 ArgProAspSerArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsn 552
Db 885 AGGCTGACTCTAGTGTTCATCATTCAGCCGGTTCCTCAATCAGATTGAGCGCTGTAAAC 944
QY 553 SerLeuAlaPheGlyGlyTyrLeuSerGluValValGluGluAsnThrProProLysMet 572
Db 945 TCITTAGCAITTTGGAGGCTACTTAAAGTGAAGTAGTAGAAGAAATATCTCCACCAAAATG 1004
QY 573 GluLysGluGlyLeuGluIleMetIleGlyLysLysLysGlyIleGlnGlyHisTyrAsn 592
Db 1005 GAAAAAGAGGCTTGGAGATAATGATTGGGAAGAAAGAAAGGATCCAGGGTCAATTACAAT 1064
QY 593 SerCysTyrLeuAspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThr 612
Db 1065 TCTTGTACTTAGACTCAACCTTATCTGCTTATTTGCTTTTGTCTGTGACACT 1124
QY 613 ValLeuLeuArgProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeu 632
Db 1125 GTGTACTTAGACCCCAAGAAAGAACGATGTAGANTATATATAGTGAACCCAGAGCTA 1184
QY 633 LeuArgThrGluIleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrLysIle 652
Db 1185 CTGAGGACAGAAATTTGTAATCTCTGAGAATAATGGAATATGTGTGTCACAAAATTT 1244
QY 653 MetLysLeuArgLysIleLeuGluLysValGluAlaSerGlyPheThrSerGluGlu 672
Db 1245 ATGAAACTGAGGAAATATCTTGAAGGTGAGGCTGCATCAGGATTTTACCTCTGAAGAA 1304
QY 673 LysAspProGluGluPheLeuAsnIleLeuPheHisIleLeuAtqValGluProLeu 692
Db 1305 AAAGATCTGAGGAATTTCTTGAATATTTCTGTTTCATCATATTTTAAGGTTAGAACCTTTG 1364
QY 693 LeuLysIleArgSerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMet 712
Db 1365 CTAAAAATAAGATCAGCAGGTCAAAGAGTACAAAGATTGTTACTTCTATCAATTTTATG 1424
QY 713 GluLysAsnGluLysValGlyValProThrIleGlnGlnLeuLeuGluTyrSerPheIle 732
Db 1425 GAAAAAATGAGAAAGTTGGCGTTCCCAATTCAGCAGTTGTTAGAAATGCTCTTTTATC 1484
QY 733 AsnSerAsnLeuLysPheAlaGluAlaProSerCysLeuIleIleGlnMetProArgPhe 752
Db 1485 AACAGTAACCTGAAATTTGAGAGGACCAATCATGTCTGATTATTCAGATCCCTCGATT 1544
QY 753 GlyLysAspPheLysLeuPheLysLysIlePheProSerLeuGluLeuAsnIleThrAsp 772
Db 1545 GGAAGAGACTTAAATCTATTTAAAAAATTTTCTCTCTGGAATTAATATAACAGAT 1604
QY 773 LeuLeuGluAspThrProArgGlnCysArgIleCysGlyGlyLeuAlaMetTyrGluCys 792
Db 1605 TTACTTGAAGACACTCCAGACAGTGCAGGATATGTGAGGCGCTTGCAATGTATGAGTGT 1664
QY 793 ArgGluCysTyrAspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThr 812
```

```
|||||
1665 AGAGAATGCTACGAGATCCGACATCTCAGCTGGAAAAATCAAGCAGTTTGTGTAACCC 1724
QY CysAsnThrGlnValHisLeuHisProLysArgLeuAsnHisLeuSerProValSer 832
Db TCACACACTCAAGTCCACCTTCATCCGAAGAGCGTGAATCATATAATATAACCCAGTGTC 1784
QY LeuProLysAspLeuProAspTrpAspTrpArgHisGlyCysIleProCysGlnAsnMet 852
Db CTTCCAAAGACTTACCGACTGGGACTGGAGACAGCGCTGCATCCCTTGCAGATATG 1844
QY GluLeuPheAlaValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGly 872
Db CAGTTATTGCTGTTCTCTGCATAGAAACAAGCCACTATGTTGCTTTGTGAAGTATGG 1904
QY LysAspAspSerAlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlnAsn 892
Db AAGGACGATTCGCTCGGCTCTCTTCGACATGCGCGATGGCGATGGTGGTGGT 1964
QY GlyPheAsnIleProGlnValThrProCysProGluValGlyGluTyrLeuLysMetSer 912
Db GGTTCACATTCCTCAAGTCAACCCATGCCAGAGTAGGAGTACTTGAAGATGCT 2024
QY LeuGluAspLeuHisSerLeuAspSerArgArgIleGlnGlyCysAlaArgArgLeuLeu 932
Db CTGGAAGACCTGCATTCCTTTGGACTCCAGGAGATCCAAAGGCTGTGCACGAAGACTGCTT 2084
QY CysAspAlaTyrMetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949
Db TGTGATGCATATATGTGCATGTACCAGATGCCAACAATGAGTTGTACAAA 2135

RESULT 13
AK000187 2341 bp mRNA linear PRI 13-SEP-2003
LOCUS Homo sapiens cDNA FLJ20180 fis, clone COL10238, highly similar to
DEFINITION AB020656 Homo sapiens mRNA for KIAA0849 protein.
ACCESSION AK000187
VERSION AK000187.1 GI:7020107
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2341)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
Minato-ku, Tokyo 108-8639, Japan (E-mail:flcdna@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction; 5'- and 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
Location/Qualifiers
1..2341
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="COL10238"
/tissue_type="colon"
/clone_lib="COL"
/note="cloning vector pME18SFL3"
```

```
misc_feature 1..2341
/note="highly similar to AB020656 Homo sapiens mRNA for
KIAA0849 protein"

ORIGIN
Alignment Scores:
Pred. No.: 2,32e-258 Length: 2341
Score: 3433.50 Matches: 644
Percent Similarity: 99.69% Conservative: 0
Best Local Similarity: 99.69% Mismatches: 1
Query Match: 68.21% Indels: 1
DB: 9 Gaps: 1
US-09-671-687A-3 (1-949) x AK000187 (1-2341)

QY 305 ThrGlnGluArgProProLysLeuAlaPheMetSerArgGlyValGlyAspLysGly 324
Db 2 ACGCAGGAAGAGAGCGCTCCCAACTTGCCTTTATGTCAAGAGGTGTTGGGACAAAGGT 61
QY 325 SerSerHisAsnLysProLysAlaThrGlySerThrSerAspProGlyAsnArg--- 343
Db 62 TCATCCAGTCATAATAAACCAAGGCTACAGGATCTACCTCAGACCCCTGGAATAGAAAC 121
QY 344 ArgSerGluLeuPheTyrThrLeuAsnGlySerSerValAspSerGlnProGlnSerLys 363
Db 122 AGATCTGAATTATTTATACCTTAATGGTCTTCTGTGTGACTCACACCAACATCCAA 181
QY 364 SerLysAsnThrTrpTyrIleAspGluValAlaGluAspProAlaLysSerLeuThrGlu 383
Db 182 TCAAAAATACATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241
QY 384 IleSerThrAspPheAspArgSerSerProProLeuGlnProProProValAsnSerLeu 403
Db 242 ATATCTACAGACTTTTGACCGCTTCTTCCACACACACCTCCAGCCCTCCTCTGTGAAC 301
QY 404 ThrThrGluAsnArgPheHisSerLeuProPheSerLeuThrLysMetProAsnThrAsn 423
Db 302 ACCACCGAGACAGATTCACCTTTTACCATTCAGTCTCACCAGATGCGCCATACCAAT 361
QY 424 GlySerIleGlyHisSerProLeuSerLeuSerAlaGlnSerValMetGluGluLeuAsn 443
Db 362 GGAAGTATTGGCCACAGTCCACTTTCTCTGTGAGCCAGCTGTGTAATGGAAGAGCTAAA 421
QY 444 ThrAlaProValGlnGluSerProProLeuAlaMetProProGlyAsnSerHisGlyLeu 463
Db 422 ACTGCACCGCTCCCAAGAGAGTCCACCTTGGCCATGCCTCTCCGGGAACATCACAATG 481
QY 464 GluValGlySerLeuAlaGluValLysGluAsnProProPheTyrGlyValIleArgTyr 483
Db 482 GNAAGTGGGCTCATTTGGCTGAAGTTAAGAGAAACCCCTCTTTCTATGGGGTAATCCG 541
QY 484 IleGlyGlnProProGlyLeuAsnGluValLeuAlaGlyLeuGluLeuGluAspGlyCys 503
Db 542 ATCCGTCAGCCACACAGGACTGAATGAAGTGTCTCGTGGACTGGAACTGGAAGATGAGT 601
QY 504 AlaGlyCysThrAspGlyThrPheArgGlyThrArgTyrPheThrCysAlaLeuLysLys 523
Db 602 GCAGGCTGTACGATGGAACCTTTCAGAGGCACCTCGGTATTTCACCTGTGCCCTGAAGA 661
QY 524 AlaLeuPheValLysLeuLysSerCysArgProAspSerArgPheAlaSerLeuGlnPro 543
Db 662 GGCCTGTTTGTGAACCTGAAGAGCTGCAGGCTGACTCTAGGTTCATTCATTCGACCG 721
QY 544 ValSerAsnGlnIleGluArgCysAsnSerLeuAlaPheGlyGlyTyrLeuSerGluVal 563
Db 722 GTTTCCAATCAGATTGAGCGCTGTAACCTTTTAGCATTTGGAGGCTACTTTAAGTGAAG 781
QY 564 ValGluGluAsnThrProLysMetGluLysGluGlyLeuGluLeuMetIleGlyLys 583
Db 782 GTAGAAGAAAAATACCTCCACCAAAAAATGGAAGAGGCTTGGAGATATATGTTGGGAAG 841
QY 584 LysLysGlyIleGlnGlyHisTyrAsnSerCysTyrLeuAspSerThrLeuPheCysLeu 603
|||||
```

```
Db 842 AAGAAAGGCATCCAGGTCATTAACAATCTTGTACTTAGACTCAACTTATTCGTGTTA 901
Qy 604 PheAlaPheSerValLeuAspThrValLeuLeuArgProLysGluLysAsnAspVal 623
Db 902 TTTGCTTTTAGTTCTGCTTCTGGACACTGTGTACTTAGACCCAAAGAAAGACGATGTA 961
Qy 624 GluTyrTyrSerGluThrGlnGluLeuLeuArgThrGluLysValLeuAsnProLeuArgIle 643
Db 962 GAATATTATAGTAAACCCAGAGCTACTGAGGACAGAAATTTGTTAATCTCTGAGAATA 1021
Qy 644 TyrGlyTyrValCysAlaThrLysIleMetLysLeuArgLysIleLeuGluLysValGlu 663
Db 1022 TAGGATATGTGTGGCCAAATAATATGAACCTGAGGAAATACCTTGAAAGAGTGGAG 1081
Qy 664 AlaAlaSerGlyPheThrSerGluGluLysAspProGluGluPheLeuAsnIleLeuPhe 683
Db 1082 GCTGCATCAGATTGCTCTGAAGAAAGATCCTCAGGAATCTTGAATATCTGTTT 1141
Qy 684 HisHisIleLeuArgValGluProLeuLeuLysIleArgSerAlaGlyGlnLysValGln 703
Db 1142 CATCATATTTTAAGGGTAGAACCTTTGCTTAAAAATAAGATCAGCAGGTCAAAAGGTACAA 1201
Qy 704 AspCysTyrPheTyrGlnIlePheMetGluLysAsnGluLysValGlyValProThrIle 723
Db 1202 GATGTACTTCTATCAAAATTTTATGAAAAAATGAGAAAGTTGGCGTTCCACAAAT 1261
Qy 724 GlnGlnLeuLeuGluTyrSerPheIleAsnSerAsnLeuLysPheAlaGluAlaProSer 743
Db 1262 CAGCAGTGTGTAGATGGTCTTTATCAACAGTAACCTGAAATTTGCAGAGGCCACCATCA 1321
Qy 744 CysLeuIleIleGlnMetProArgPheGlyLysAspPheLysLeuPheLysIlePhe 763
Db 1322 TGTCTGATTATTCAGATGCTCGATTTGGAAAAAGACTTTAAACTATTTAAAAAATTTT 1381
Qy 764 ProSerLeuLeuLeuAsnIleThrAspLeuLeuGluAspThrProArgGlnCysArgIle 783
Db 1382 CCTTCTCTGGAATTAATAATAACAGATTTACTTTGAAGACACTCCACAGACAGTCGCGATA 1441
Qy 784 CysGlyGlyLeuAlaMetTyrGluCysArgGluCysTyrAspAspProAspIleSerAla 803
Db 1442 TGTGGAGGGCTTCCAATGTATGAGTGAGTAAGATGCTTACCAGCATCCGGAATCTCAGCT 1501
Qy 804 GlyLysIleLysGlnPheCysLysThrCysAsnThrGlnValHisLeuHisProLysArg 823
Db 1502 GGAATAATCAAGCAGTTTGTAAACCTTGCAACACTCAAGTCCACCTTCATCCGAGAGG 1561
Qy 824 LeuAsnHisLysTyrAsnProValSerLeuProLysAspLeuProAspThrAspTrpArg 843
Db 1562 CTGAATCATAAATAATAACCCAGTGTCACTTCCCAAGACTTACCAGACTGGGACTGGAGA 1621
Qy 844 HisGlyCysIleProCysGlnAsnMetGluLeuPheAlaValLeuCysIleGluThrSer 863
Db 1622 CAGGCTGCATCCCTTGCCAGATAATGAGTTATTTGCTGTTCTCTGCATAGAAACAAGC 1681
Qy 864 HisTyrValAlaPheValLysTyrGlyLysAspAspSerAlaTrpLeuPhePheAspSer 883
Db 1682 CACTATGTGCTTTGTGAAGTATGGAGACGATTTCTGCCCTGGCTCTCTTTTGACAGC 1741
Qy 884 MetAlaAspArgAspGlyGlyGlnAsnGlyPheAsnIleProGlnValThrProCysPro 903
Db 1742 ATGGCCGATCGGATGGTGGTCAGATGGCTTCAACATTCCTCAAGTCAACCCCATGCCCA 1801
Qy 904 GluValGlyGluTyrLeuLysMetSerLeuGluAspLeuHisSerLeuAspSerArgArg 923
Db 1802 GAAGTAGGAGAGTACTTGAAGATGTCTCTGGAAGACCTGCAATTCCTTGGACTCCAGAGA 1861
Qy 924 IleGlnGlyCysAlaArgLeuLeuCysAspAlaTyrMetCysMetTyrGlnSerPro 943
Db 1862 ATCCAAGGCTGTGCACGAGACTGCTTTGTGATGCATATATGTCATGTCACGAGTCCA 1921
Qy 944 ThrMetSerLeuTyrLys 949
Db 1922 ACATCAGTTTGTACAAA 1939
```

RESULT 14

```
AX883681 LOCUS 2569 bp DNA linear PAT 17-DEC-2003
Sequence 18586 from Patent EP1074617.
ACCESSION AX883681
VERSION AX883681.1 GI:40038582
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota, T., Isozaki, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE Primers for synthesizing full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 18586, 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
source
1. 2569
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
403..1833
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAE92065.1"
/db_xref="GI:40038583"
/translation="MQVELPPLEINSRVSLKVGETIESGTVIFCDVLPKESLSGVFVG
VMDNPIGNWDGRFQVQLCSFACVESTILLHINDIIPESVTQRRPKLAFMRGVRG
DKSSSHNPKATGSDTPGNRNSLFTYTLNGSSVDSPQSKNTWTYIDVADPA
KSLTEITDTPDRSPPLQPPVNSLTTFNPSLPLSLTKMPTNTSGTGHSPLSAQ
SVMEELNTAPQESPPPLAMPNGSHGLEVSLAEVKENPPPVGVIRWIGQPPGLNEVL
AGLEDEACAGCTDGTFRGTYRTFCAKALFVKLSCRPSRSPASQPVSNQLERN
SLAFGLYLSVEVEENTPPRMKEGLEIMIGKKGIQGHYNSCYLDSTLFCIFAFSSVL
DTLLRPKNDVEYSETQELLRTIVNPLRIYGVYCATKIMKLRLKLEKVEAASGF
TSEKGDHLNLALKITEIHMSKY"
ORIGIN
```

Alignment Scores:

```
Pred. No.: 3 52e-226 Length: 2569
Score: 3022.50 Matches: 587
Percent Similarity: 98.66% Conservative: 2
Best Local Similarity: 98.32% Mismatches: 3
Query Match: 60.04% Indels: 5
DB: 6 Gaps: 4
```

US-09-671-687A-3 (1-949) x AX883681 (1-2569)

```
Qy 88 ValValGluIleAsnGluLysPheThrGluLeuLeuAlaIleThrAsnCysGluGlu 107
Db 1 GTTGTAGAGATAAATAAGAAAGTTACACAGAGTTACTTTTGGCAATTACCAATTTGGAGAG 60
Qy 108 ArgPheSerLeuPheLysAsnArgAsnArgLeuSerLysGlyLeuGlnIleAspValGly 127
Db 61 AGTTTCAGCTGTTTAAACACAGACAGACTAAGTAAAGCCCTCCAAATAGACGTGGGC 120
Qy 128 CysProValLysValGlnLeuArgSerGlyGluLysPheProGlyValValArgPhe 147
Db 121 TGTCTCTGTAAGATACAGCTGAGATCTGGGGAAGAAAATTTCTCGAGTTGTACGCTTC 180
Qy 148 ArgGlyProLeuLeuAlaGluArgThrValSerGlyIlePhePheGlyValGluLeuLeu 167
Db 181 AGAGGACCCCTGTTAGCAGAGAGACAGCTCCCGAATATTTCTTGGAGTTGAATTCGTG 240
Qy 168 GluGluGlyArgGlyGlnGlyPheThrAspGlyValTyrGlnGlyLysGlnLeuPheGln 187
Db 241 GAAGAGGTCGTGCTCAAGGTTTCACTGACGGGTGTACCAGGGAACACCTTTTTCAG 300
Qy 188 CysAspGluAspCysGly----PheValAlaLeuAspLysLeuGluLeuIleGluAspAsp 206
Db 301 TGTGATGAAGATTGTGGCGTGTGTTGTCATTTGGCATTTGGCAAGCTAGAACTCATAGAAGATGAT 360
```

Qy	207	AspThrAlaLeuGluSerAspTyrAlaGlyProGlyAspThrMetGlnValGluLeuPro	226						
Db	361	GACACTGCATTGGAAAGTGATTACGAGGTCTCTGGGACAAATGCAGTCCAACTTCCT	420						
Qy	227	ProLeuGluIleAsnSerArgValSerLeuLysGlyGlyGluThrIleGluSerGlyThr	246						
Db	421	CCTTTGGAAATAAACTCCAGAGTTCTTTTGAAGGTGGAGAAACAATAGAAATCTGGAACA	480						
Qy	247	VallilePheCysAspValLeuProGlyLysGluSerLeuGlyTyrPheValGlyValAsp	266						
Db	481	GTTTATATCTGTGATGTTTTCGCCAGGAAAGAAAGCTTTAGGATATTTTGTGGTGTGGAC	540						
Qy	267	MetAspAsnProIleGlyAsnTrpAspGlyVargPheAspGlyVal----LeuCysSerPhe	285						
Db	541	ATGGATAACCCCTATTGGCAACTGGGATGGAGATTGTGAGGTGCAGCTTTGTAGTTTTT	600						
Qy	286	AlaCysValGluSerThrIleLeuLeuHisIleAsnAspIleIleProGluSerValThr	305						
Db	601	CGGTGTGTGGAAGTACAAATCTATTGCACATCAATGATATCATCCACAGAGAGTGTGACG	660						
Qy	306	GlnGluArgArgProProLysLeuAlaPheMetSerArgGlyValGlyAspLysGlySer	325						
Db	661	CAGGAAAGGAGCCCTCCCAAACTTGCCTTTATGTCAAGAGGTGTGGGACAAAGGTTCA	720						
Qy	326	SerSerHisAsnLysProLysAlaThrGlySerThrSerAspProGlyAsnArg---Arg	344						
Db	721	TCCAGTCAATAATAACCAAGAGCTACAGGATCTACCTCAGACCCCTGGAAATAGAAACAGA	780						
Qy	345	SerGluLeuPheTyrThrLeuAsnGlySerSerValAspSerGlnProGlnSerLysSer	364						
Db	781	TCTGAATTTATTTATACCTTAATGGGTCTCTGTGTGACTCACAAACCAACAATCCAAATCA	840						
Qy	365	LysAsnThrTrpTyrIleAspGluValAlaGluAspProAlaLysSerLeuThrGluIle	384						
Db	841	AAAAATACATGATGATGATGAGTTCAGAGAGACCCCTGCAGAAATCTCTTACAGAGATA	900						
Qy	385	SerThrAspPheAspArgSerSerProLeuGlnProProProValAsnSerLeuThr	404						
Db	901	TCTACAGACTTTGACCGTCTTCCACCACCTCCAGCCCTCCCTCGTGAACCTCAGCTACC	960						
Qy	405	ThrGluAsnArgPheHisSerLeuProPheSerLeuThrLysMetProAsnThrAsnGly	424						
Db	961	ACCGAGAACAGATTCATCTTTTACCATTCAGTCTCACCAAGATGCCCAATACCAATGGA	1020						
Qy	425	SerIleGlyHisSerProLeuSerLeuSerAlaGlnSerValMetGluGluLeuAsnThr	444						
Db	1021	AGTATTTGCCACAGTCCACTTTCTCTGTGAGCCAGTCTGTAATGGAAGAGCTAAACACT	1080						
Qy	445	AlaProValGlnGluSerProProLeuAlaMetProProGlyAsnSerHisGlyLeuGlu	464						
Db	1081	GCACCCGTCCAAGAGAGTCCACCTTTGGCCATGCTCTCTGGGAACCTCACATGGTCTAGAA	1140						
Qy	465	ValGlySerLeuAlaGluValLysGluAsnProProPheTyrGlyValIleArgTrpIle	484						
Db	1141	GTGGGCTCATTTGGCTGAAAGTAAAGAGAACCCCTCTCTTATGGGGTAAATCCGTTGGATC	1200						
Qy	485	GlyGlnProProGlyLeuAsnGluValLeuAlaGlyLeuGluLeuGluAspGluCysAla	504						
Db	1201	GGTCAGCCACCAAGGACTGAATGAAGTGTCTGCTGGACTGGAACTGGAAGATGAGTGTGCA	1260						
Qy	505	GlyCysThrAspGlyThrPheArgGlyThrArgTyrPheThrCysAlaLeuLysLysAla	524						
Db	1261	GGCTGTACGGATGGAACCTTCAGAGGCACTCGGTATTTTCCCTGTGTGCCCTGAAGAAGCG	1320						
Qy	525	LeuPheValLysLeuLysSerCysArgProAspSerArgPheAlaSerLeuGlnProVal	544						
Db	1321	CTGTTTGTGAAACTGAAGAGCTGCAGGCGTCACTCTAGGTTTGCATCTTGAGCCGGT	1380						
Qy	545	SerAsnGlnIleGluArgCysAsnSerLeuAlaPheGlyGlyTyrLeuSerGluValVal	564						
Db	1381	TCCAAATCAGATTGAGCGCTGTAACTCTTTAGCATTTGGAGGCTACTTTAAGTGAAGTAGTA	1440						
Qy	565	GluGluAsnThrProProLysMetGluLysGluGlyLeuGluIleMetIleGlyLysLys	584						
Db									
Db	1441	GAAGAAATACTCCACCAAAATATGAAAAGAGGCTTGGAGATAATGATTGGGAAGAG	1500						
Qy	585	LysGlyIleGlnGlyHisTyrAsnSerCysTyrLeuAspSerThrLeuPheCysLeuPhe	604						
Db	1501	AAAGGCATCCAGGTCAATTACAAATCTTGTACTTAGACTCAACCTTATCTTCCTTATTT	1560						
Qy	605	AlaPheSerSerValLeuAspThrValLeuLeuArgProLysGluLysAsnAspValGlu	624						
Db	1561	GCTTTTAGTCTCTGTGGACACTGTGTACTTACCCCAAGAAAGAACCATGTAGAA	1620						
Qy	625	TyrTyrSerGluThrGlnGluLeuLeuArgThrGluIleValAsnProLeuArgIleTyr	644						
Db	1621	TATTATAGTGAACCCAGAGCTACTGAGGACAGAAATGTTAATCTCTCGAATAATAT	1680						
Qy	645	GlyTyrValCysAlaThrLysIleMetLysLeuArgLysIleLeuGluLysValGluAla	664						
Db	1681	GGATATGTGTGTGCCACAAAATATGAACTAGGAAAAATACTTGAAAAGGTGGAGGCT	1740						
Qy	665	AlaSerGlyPheThrSerGluGluLysAspProGluGluPheLeuAsnIle	681						
Db	1741	GCATCAGGATTTACCTCTGAAGAAAAAGGT-----GACCATCTTAACTTA	1785						
RESULT 15									
LOCUS	BD160470	2569 bp	DNA	linear					
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.								
ACCESSION	BD160470								
VERSION	BD160470.1	GI:27866228							
KEYWORDS	JP 2002191363-A/15313.								
SOURCE	Homo sapiens (human)								
ORGANISM	Homo sapiens								
REFERENCE	1 (bases 1 to 2569)								
AUTHORS	Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.								
TITLE	Primer for synthesizing full-length cDNA and use thereof								
JOURNAL	Patent: JP 2002191363-A 15313 09-JUL-2002;								
COMMENT	HELIIX RESEARCH INSTITUTE OS Homo sapiens (human) PN JP 2002191363-A/15313 PD 09-JUL-2002 PF 28-JUL-2000 JP 2000280990 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO, PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU, PI KEIICHI NAGAI, TETSUJI OTSUKI PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC Primer for synthesizing full-length cDNA and use thereof FH Key								
FEATURES	FT	CDS	Location/Qualifiers (403).. (1830).						
source	1..2569 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"								
ORIGIN									
Alignment Scores:	3,52e-226	Length:	2569						
Pred. No.:	3022.50	Matches:	587						
Score:	98.66%	Conservative:	2						
Percent Similarity:	98.32%	Mismatches:	3						
Best Local Similarity:	60.04%	Indels:	5						
Query Match:	6	Gaps:	4						
DB:									
US-09-671-687A-3 (1-949) x BD160470 (1-2569)									
Qy	88	ValValGluIleAsnGluLysPheThrGluLeuLeuAlaIleThrAsnCysGluGlu	107						

```
Db      1 GTTGTAGAGATAAATGAAAGTTCCACAGATTACTTTGGCAATTACCAATTTGTGAGGAG 60
Qy      ArgPheSerLeuPheLysAsnArgAsnArgLeuSerLysGlyLeuGlnIleAspValGly 127
Db      61 AGGTTACAGCTGTTTAAAACAGAAAACAGACTAAGTAAGGCCTCCAAANTAGACGTGGGC 120
Qy      128 CysProValLysValGlnLeuArgSerGlyGluGlyLysPheProGlyValValArgPhe 147
Db      121 TGTCTCTGAAAGTACAGCTGAGACTCTGGGGAAGAAAAATTTCTCGAGTTGTACGCTTC 180
Qy      148 ArgGlyProLeuLeuAlaGluArgThrValSerGlyIlePhePheGlyValGluLeuLeu 167
Db      181 AGAGGACCCCTGTTAGCAGAGAGACAGTCTCCGGAATATTCTTTGGAGTTGAATTTGCTG 240
Qy      168 GluGluGlyArgGlyGlnGlyPheThrAspGlyValTyrGlnGlyLysGlnLeuPheGln 187
Db      241 GAAGNAGTCTGTTGCTCAAGGTTTCTACTGACGGGTGTACNAGGGAACACAGCTTTTCAG 300
Qy      188 CysAspGluAspCysGly---PheValAlaLeuAspLysLeuGluLeuIleGluAspAsp 206
Db      301 TGTGATCAAGATTGTGGCGTGTGTTGTCATTGGCAAGCTAGAACTCATAGAAGATGAT 360
Qy      207 AspThrAlaLeuGluSerAspTyrAlaGlyProGlyAspThrMetGlnValGluLeuPro 226
Db      361 GACACTGCATTGGAAGTGATTAACGACAGTCTCGGGACACAAATGCAGGTTCGAACCTTCT 420
Qy      227 ProLeuGluIleAsnSerArgValSerLeuLysGlyGlyGluThrIleGluSerGlyThr 246
Db      421 CCTTTGGAAATAAAGCTCCAGAGTTTCTTTGAAGGTTGGGAAACAAATAGAACTTCGAAACA 480
Qy      247 ValIlePheCysAspValLeuProGlyLysGluSerLeuGlyTyrPheValGlyValAsp 266
Db      481 GTTATATTCTGTGATGTTTTCGCCAGGAAAGAAAGCTTAGGATATTTTGTGGTGGAC 540
Qy      267 MetAspAsnProIleGlyAsnTrpAspGlyArgPheAspGlyVal---LeuCysSerPhe 285
Db      541 ATGGATAAACCTATTGGCAACTGGGATGGAAGATTTTGTGAGGTGACGCTTTGTAGTTT 600
Qy      286 AlaCysValGluSerThrIleLeuLeuHisIleAsnAspIleIleProGluSerValThr 305
Db      601 GCCTGTGTTGAAAGTCAATTCATTTCACATCAATGATATATCCTCCAGAGAGTGTGACG 660
Qy      306 GlnGluArgArgProProLysLeuAlaPheMetSerArgGlyValGlyAspLysGlySer 325
Db      661 CAGGAAGAGGGGCTCCCAAACCTTGCTTTATGTCAAGAGGTGTTGGGGACAAAGGTICA 720
Qy      326 SerSerHisAsnLysProLysAlaThrGlySerThrSerAspProGlyAsnArg---Arg 344
Db      721 TCCAGTCATAATAAACCAAGGCTTACAGGATCTACCTCAGACCTCGGAATAGAAACAGA 780
Qy      345 SerGluLeuPheTyrThrLeuAsnGlySerSerValAspSerGlnProGlnSerLysSer 364
Db      781 TCTGAATATATTTTATACCTTTAAATGGGTCTTCTGTTGACTCACAAACCACAAATCA 840
Qy      365 LysAsnThrTrpTyrIleAspGluValAlaGluAspProAlaLysSerLeuThrGluIle 384
Db      841 AAAAAATACATGGTACATTTGATGAAGTTGCAAGAGACCTCGCAAAATCTCTTACAGAGATA 900
Qy      385 SerThrAspPheAspArgSerSerProProLeuGlnProProValAsnSerLeuThr 404
Db      901 TCTACAGACTTTGACCGTCTCTTCCACCACCACTCCAGCCTCTCTCTGTGAACTCACTGACC 960
Qy      405 ThrGluAsnArgPheHisSerLeuProPheSerLeuThrLysMetProAsnThrAsnGly 424
Db      961 ACCGAGAACAGATTCACCTCTTTACCATTCAGTCTCACCAAGATGCCCAATACCAATGGA 1020
Qy      425 SerIleGlyHisSerProLeuSerLeuSerAlaGlnSerValMetGluGluLeuAsnThr 444
Db      1021 AGTATTTGGCCACAGTCCACCTTTCTCTGTTCAGCCAGTCTCTTAATGGGAAGAGCTAAACACT 1080
Qy      445 AlaProValGlnGluSerProProLeuAlaMetProProGlyAsnSerHisGlyLeuGlu
```

```
Db      1081 GCACCCGTCCAGAGAGTCCACCCCTTGGCCATGCTCTCTGGAACTCATCGTGTAGAA 1140
Qy      465 ValGlySerLeuAlaGluValLysGluAsnProProPheTyrGlyValIleArgTrpIle 484
Db      1141 GTGGGCTCATTTGGCTGAAGTTAAGAGAACCTCTCTTCTATGGGGTAAATCCGTTGGATC 1200
Qy      485 GlyGlnProProGlyLeuAsnGluValLeuAlaGlyLeuGluLeuGluAspGluCysAla 504
Db      1201 GGTCAAGCACACAGGACTGAATGAAGTGTCTCGCTGGACTGGAACCTGGAAGATGAGTGTGCA 1260
Qy      505 GlyCysThrAspGlyThrPheArgGlyThrArgTyrPheThrCysAlaLeuLysLysAla 524
Db      1261 GGCTGTACGGATGAAACCTTCAGAGGCACTCGGTATTTTTCACCTGTGCCCTGAAAGAAGCG 1320
Qy      525 LeuPheValLysLeuLysSerCysArgProAspSerArgPheAlaSerLeuGlnProVal 544
Db      1321 CTGTTTGTGAAACTGAAGAGCTGCAGGCTGACTCTAGGTTTGCATCATTCGACGCCGTT 1380
Qy      545 SerAsnGlnIleGluArgCysAsnSerLeuAlaPheGlyGlyTyrLeuSerGluValVal 564
Db      1381 TCCAATCAGATTGAGCGCTGTAACTCTTTAGCATTTGGAGGCTACTTAAGTGAAGTAGTA 1440
Qy      565 GluGluAsnThrProProLysMetGluLysGluGlyLeuGluIleMetIleGlyLysLys 584
Db      1441 GAAAGAAATACTCCACCAAAATGGAAGAAAGGCTTGGAGATAATGATTTGGGAAGAG 1500
Qy      585 LysGlyIleGlnGlyHisTyrAsnSerCysTyrLeuAspSerThrLeuPheCysLeuPhe 604
Db      1501 AAAGGCATCCAGGTCATTACATCTTGTGTACTTTAGACTCAACCTTAATTCCTCTTATTT 1560
Qy      605 AlaPheSerSerValLeuAspThrValLeuLeuArgProLysGluLysAsnAspValGlu 624
Db      1561 GCITTTAGTTCTGTTCTGGACACTGTGTACTTAGACCCAAAGAAAGACGATGTAGAA 1620
Qy      625 TyrTyrSerGluThrGlnGluLeuLeuArgThrGluIleValAsnProLeuArgIleTyr 644
Db      1621 TATTATAGTGAACCCCAAGAGCTACTGAGGACAGAAATTTGTTAATCTCTCAGAAATATAT 1680
Qy      645 GlyTyrValCysAlaThrLysIleMetLysLeuArgIlyIleLeuGluLysValGluAla 664
Db      1681 GGATATGTGTGTGCCCAAAATTTATGAACTGAGGAAAAATACTTGAAGAGGTGGAGGCT 1740
Qy      665 AlaSerGlyPheThrSerGluGluLysAspProGluGluPheLeuAsnIle 681
Db      1741 GCATCAGATTTACCTCTGAGAAAAAAGGT-----GACCATCTTAACTTA 1785
```

Search completed: September 18, 2004, 19:01:03
Job time : 8564 secs